

Teacher Ratings of Attention Deficit Hyperactivity Disorder Symptoms : Factor Structure and Normative Data*

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

DuPaul, G.J., Power, T.J., [Anastopoulos, A.D.](#), Reid, R., McGoey, K.E. & Ikeda, M.J. (1997). Teacher ratings of Attention-Deficit/Hyperactivity Disorder symptoms: Factor structure and normative data. *Psychological Assessment*, 9, 436-444. DOI: 10.1037/1040-3590.9.4.436

Made available courtesy of American Psychological Association:
<http://www.apa.org/journals/pas/>

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

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

Attention deficit hyperactivity disorder (ADHD) involves the display of developmentally inappropriate levels of inattention, impulsivity, and overactivity resulting in functional impairment across two or more settings (American Psychiatric Association [APA], 1994). Approximately 4% of school-aged children and adolescents in the United States have this disorder. Children with this disorder are at higher than average risk for academic underachievement, conduct problems, and social relationship difficulties, as a function of the core symptoms of ADHD (Barkley, 1990). The disorder is chronic for most individuals with ADHD and requires long-term treatment (Weiss & Hechtman, 1993). Given the chronic nature of ADHD and associated problematic outcomes, assessment measures are needed for screening and identification purposes.

An evaluation of ADHD typically includes diagnostic interviews with the child and his or her parents and teachers, behavior rating scales completed by parents and teachers, direct observations of school behavior, and clinic-based testing (Barkley, 1990; DuPaul & Stoner, 1994). A number of teacher questionnaires have been developed for screening and identification of children with ADHD, including the Conners Rating Scales (Conners, 1989), the Attention Deficit Disorder with Hyperactivity (ADD-H) Comprehensive Teacher Rating Scale (Ullmann, Sleator, & Sprague, 1985), the Attention Deficit Disorders Evaluation Scale (McCarney, 1989), the Child Attention Profile (Barkley, 1991), and the Disruptive Behavior Disorders Rating Scale (Pelham, Evans, Gnagy, & Greenslade, 1992). In addition, the ADHD Rating Scale (DuPaul, 1991) was developed to specifically obtain teacher ratings of the frequency of *Diagnostic and*

* The completion of this investigation was supported, in part, by a Faculty Research Grant from Lehigh University. We greatly appreciate the assistance of the following individuals in obtaining teacher ratings: Michelle Beck, Michael Booher, Allison Costabile, Suzanne Cote, Ralph Daubert, Joanna Gabris, Vincent Glaub, Charlene Jennings, Daniel Martin, Mary Beth Noll, Nick Novak, Joe Olmi, Daniel Reschly, Cynthia Riccio, Maura Roberts, Diana Rogers-Adkinson, James Stumme, and Vincent Rutland.

Statistical Manual of Mental Disorders (3rd ed., rev.) (*DSM-III-R*; APA, 1987) symptoms of ADHD.

Although these measures have demonstrated reliability and validity, their clinical use is limited given the change in diagnostic criteria for ADHD with the publication of the *DSM* (4th ed., *DSM-IV*; APA, 1994). The core characteristics of ADHD (i.e., inattention, impulsivity, and hyperactivity) have been retained; however, the specific symptoms and dimensional structure have been altered to a significant degree. Specifically, the *DSM-III-R* contained 14 symptoms conceptualized as comprising a unitary dimension, whereas the *DSM-IV* includes two dimensions comprised of 9 symptoms of inattention and 9 symptoms of hyperactivity–impulsivity (Lahey et al., 1994). This two-dimensional structure of ADHD symptomatology is in keeping with theoretical and conceptual views of this disorder (e.g., Barkley, 1997; Lahey & Carlson, 1992). Factor analytic work conducted with samples from the United States (DuPaul, 1991; Lahey et al., 1988), Brazil (Brito, Pinto, & Lins, 1995), Germany (Baumgaertel, Wolraich, & Dietrich, 1995), and Puerto Rico (Bauermeister et al., 1995) has provided empirical support for this two-dimensional conceptualization of ADHD. Furthermore, these two symptomatic dimensions are clinically important because they are used to define the three subtypes of ADHD: the combined type, predominantly inattentive type, and predominantly hyperactive–impulsive type. Thus, behavior questionnaires developed on the basis of the *DSM-III-R* (e.g., the ADHD Rating Scale and the Disruptive Behavior Disorders Rating Scale) are limited in clinical use for screening and identification purposes.

Another limitation of most currently available ADHD behavior rating scales is that their factorial structure has been determined using principal component analysis with varimax rotation (Reid, 1995). Although commonly used, principal component analysis is not appropriate for this purpose because this analysis creates uncorrelated linear combinations of observed variables to explain as much of the total sample variance as possible (Taylor & Sandberg, 1984).

Furthermore, the use of varimax rotation is problematic for two reasons (Gorsuch, 1983). First, varimax rotation implies that factors are orthogonal or conceptually unrelated; however, this is a questionable assumption in the case of ADHD ratings. Studies have reported high correlations between hyperactivity and inattention factor scores (e.g., McCarney, 1989; Shaywitz, Shaywitz, Schnell, & Towle, 1988; Ullmann et al., 1985). Second, because the varimax procedure is designed to maximize variance among all factors, it mitigates against producing a general factor solution and may create several “splinter” factors that could be more parsimoniously combined. In scales with high-internal consistency, data should be obliquely rotated (Gorsuch, 1983).

The purpose of this study was to develop a rating scale containing *DSM-IV* criteria for ADHD that could be completed by teachers in the context of screening, multimethod assessment, or both of this disorder. A large, national sample was obtained to ascertain (a) whether teacher ratings of ADHD symptoms in a community sample conformed to the two-factor model delineated in the *DSM-IV*, using exploratory and confirmatory factor analyses with oblique rotation, and (b) whether the frequencies of ADHD symptoms varied as a function of sex, age, and ethnic group. Normative data were obtained using a subsample of children as representative of the 1990 U.S. census distributions for region and ethnic group.

Method

Participants

Two samples of children were used in this study. The second sample was a subset of the first sample. Sample 1 was composed of 4,130 children and adolescents between 4 and 19 years old from 31 school districts across the United States. This sample was used for factor analyses and examination of the effects of sex, age, and ethnic group on ADHD ratings. Complete ratings of ADHD symptoms were available for 4,009 participants (2,054 boys, 1,934 girls, and 21 unspecified). Students ranged in age from 4 to 19 years old ($M = 10.3$, $SD = 3.5$) and attended kindergarten through 12th grade ($M = 4.8$, $SD = 3.5$). Participants were predominantly Caucasian ($n = 2,785$, 69.5%), with additional students identified as African American ($n = 735$, 18.3%), Hispanic ($n = 229$, 5.7%), Asian American ($n = 74$, 1.8%), Native American ($n = 4$, 0.1%), other ($n = 119$, 3.0%), and unspecified ($n = 63$, 1.6%). Most participants attended general education ($n = 3,612$) versus special education ($n = 336$) classrooms. A total of 2,005 teachers (1,605 female, 371 male, and 29 unspecified) participated, with most teachers completing ratings on one boy and one girl in their classroom. As was the case for child participants, most teachers were Caucasian ($n = 1,817$, 90.6%), with additional teachers identifying themselves as African American ($n = 113$, 5.6%), Hispanic ($n = 25$, 1.2%), Asian American ($n = 5$, 0.2%), Native American ($n = 1$, < 0.1%), other ($n = 13$, 0.6%), and unspecified ($n = 31$, 1.5%). There was a wide range in years of teaching experience (0–44 years, $M = 14.6$, $SD = 9.24$).

Sample 2 was composed of 2,000 (1,040 boys, 948 girls, and 12 unspecified) randomly selected participants from Sample 1. As described below, this sample was selected to approximate the 1990 U.S. census data distributions for ethnic group and region. Sample 2 was used to derive normative data that could be used for screening and identifying purposes. Participants ranged in age from 4 to 19 years old ($M = 10.6$, $SD = 3.6$) and attended kindergarten through 12th grade ($M = 5.1$, $SD = 3.5$). Most children attended general education ($n = 1,816$) versus special education ($n = 161$) classrooms. The racial distribution of the sample was 65.1% Caucasian ($n = 1,303$), 18.5% African American ($n = 369$), 8.0% Hispanic ($n = 160$), 1.7% Asian American ($n = 34$), 0.2% Native American ($n = 3$), and 5.2% other ($n = 104$), with an additional 27 participants (1.4%) of unspecified ethnic origin. Participants were identified as living in one of four regions of the United States, with 34.0% ($n = 680$) from the South, 29.5% ($n = 590$) from the Midwest, 20.0% ($n = 400$) from the Northeast, and 16.5% ($n = 330$) from the West. A total of 1,001 teachers (793 female, 194 male, and 14 unspecified) completed ratings. Teachers were predominantly Caucasian ($n = 902$, 90.2%), whereas the others identified themselves as African American ($n = 61$, 6.1%), Hispanic ($n = 13$, 1.3%), Asian American ($n = 3$, 0.3%), Native American ($n = 1$, 0.1%), other ($n = 6$, 0.6%), or of unspecified ethnic background ($n = 15$, 1.5%). As was the case with Sample 1, a wide range of teaching experience was reported (0–44 years, $M = 14.3$, $SD = 9.1$).

Measures

Teachers for both samples were asked to provide information regarding their sex, ethnic group, and years of teaching experience as well as the type of classroom (general vs. special education) and grade level that they taught. Information about the child to rate, such as sex, ethnic group, and age, also was provided.

Teachers also completed the ADHD Rating Scale–IV (school version, ARS; DuPaul, Anastopoulos, Power, Murphy, & Barkley, 1994; see the Appendix), which consists of 18 items

directly adapted from the ADHD symptom list as specified in the *DSM-IV*. To address possible response bias, we designated inattention symptoms as odd-numbered items and hyperactivity–impulsivity symptoms 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 four-point Likert scale, ranging from *never or rarely* (0) to *very often* (3), with higher scores indicative of greater ADHD-related behavior. ARS scores have been found (a) to have adequate levels of internal consistency and stability over a 4-week period and (b) to be correlated significantly with direct observations of classroom behavior (DuPaul, Power, McGoey, Ikeda, & Anastopoulos, 1997).

Procedures

Sample 1 was obtained by asking teachers in the 31 districts to complete an ARS with regard to the performance of two randomly selected students (one boy and one girl) from their classroom. Because ratings were anonymous and confidential, informed consent from parents was not obtained. Teachers were asked to rate the behavior of students located at different points on their class roster (e.g., fifth girl and eighth boy). Teachers at the secondary level, who taught more than one class, were asked to rate two randomly selected students from a randomly selected class (e.g., third period). For districts in suburban–rural locations ($n = 22$), all teachers were asked to participate. In the larger urban districts ($n = 9$), schools at each level were selected to provide a cross-section of the community and all teachers at selected schools were invited to participate. Ratings were completed between October and May in either the 1994–1995 or 1995–1996 school year, with estimated return rates ranging from 50% to 95% ($M = 85%$) across school districts.

Normative data (i.e., Sample 2) were obtained by selecting a subsample of ratings from Sample 1 to conform, as closely as possible, to U.S. census data population proportions regarding region and ethnic group. Participants were randomly selected from Sample 1 in a stratified manner (i.e., random selection was constrained to conform with proportions of regional and ethnic distribution). The resulting normative sample closely matched U.S. census distributions for region with the exception of the Western part of the United States, which was somewhat underrepresented in the sample (i.e., 16.5% in the sample vs. 21.0% in the census). African Americans were slightly overrepresented (18.5% in the sample vs. 12.0% in the census) in the normative group. Nevertheless, differences between sample proportions and U.S. census data are minimal; thus we considered Sample 2 to be an adequate, representative sample of the U.S. child population.

Results

Two types of factor analyses were conducted. First, an exploratory factor analysis was conducted using principal axis factoring (PAF) and oblique rotation, which allows for factors to correlate. Two solutions were computed: a forced one-factor solution that served as a baseline and an unforced solution that extracted all factors with eigenvalues greater than 1.00. Second, linear structural equation modeling was used to perform a confirmatory factor analysis (CFA) to test the results of the PAF solution.

Exploratory Factor Analysis

Analysis was performed on data from Sample 1 ($N = 4,009$) for which ratings on all 18 items were available. Visual analysis of the correlation and anti-image correlation (i.e., the negative of the partial correlation coefficients) matrices and the results of the Kaiser–Meyer–Olkin measure of sampling adequacy (.97) indicated that the data were adequate for analysis. Table 1 shows the factor loadings and communalities for the one-factor and unforced solutions. The forced one-factor solution accounted for 64.8% of the variance. As Table 1 indicates, all item loadings were in the .7 to .8 range. The unforced analysis resulted in a two-factor solution that accounted for 71.9% of the variance. Eigenvalues (and percentage of variance explained) for Factors 1 and 2 respectively were 11.38 (63.2%) and 1.57 (8.7%). After factor rotation, the proportion of explained variance accounted for by each factor was calculated by summing the squared loadings for each factor and dividing by the number of items (Gorsuch, 1983). Each of the two factors accounted for 33.0% of explained variance. The sum is less than the 71.9% total because factors were allowed to correlate and the loadings reflect only unique variance. The pattern matrix approximates simple structure, with even-numbered items (i.e., those reflecting hyperactivity–impulsivity) loaded on Factor 1 and odd-numbered items (i.e., reflecting inattention) loaded on Factor 2. Only items 2 and 5 loaded on both factors; however, in both cases, loadings were relatively low ($< .40$) on the secondary factor. The correlation between the two factors was $-.70$. The interpretation of the between-factor correlation should be on the basis of the absolute magnitude of this statistic. The negative correlation between factors is most likely an artifact related to the nature of the inattention factor (i.e., higher scores represent less attention) and that with a simple structure there is an inverse relation between factors.

Table 1
Rotated Pattern Matrix for Teacher Ratings

Item	Forced one-factor solution		Two-factor solution		
	Communality	Loading	Communality	Factor 1	Factor 2
14	.51395	.71690	.70173	.93235	.14384
18	.67337	.82059	.82190	.92376	.02472
16	.64720	.80449	.79978	.92227	.04051
10	.59497	.77134	.69933	.82345	-.01812
12	.56392	.75094	.65439	.78556	-.03285
6	.54949	.74127	.61623	.72938	-.07655
4	.65826	.81155	.70977	.72826	-.15270
8	.63222	.79512	.66390	.66339	-.19818
2	.67876	.82272	.68186	.56983	-.31917
7	.60135	.77546	.78576	-.11140	-.96105
9	.65078	.80671	.79478	-.03427	-.91522
11	.60234	.77611	.73550	-.03509	-.88187
17	.60135	.78963	.75199	-.01957	-.88079
1	.58827	.76699	.69768	-.00052	-.83564
3	.71569	.84599	.76067	.17971	-.73660
13	.60568	.77825	.66263	.11365	-.73022
15	.76944	.87718	.78440	.29200	-.65596
5	.62186	.78858	.62628	.30040	-.55108

Some items (e.g., items 7, 9, 14, 16, and 18) exhibited extremely high loadings (i.e., $> .90$). For loadings of this magnitude, it is difficult to distinguish the item from the underlying factor itself, which may distort factor structure. To determine if this had occurred, we conducted additional exploratory analyses using the same procedures, but in each analysis one of the items with a high loading was excluded. Exclusion of items did not result in any discernible change in the factor structure or great disparity in the loading of other items. Subsequent analyses showed that it was

possible to exclude both items 14 and 18 without any appreciable loss in explained variance or alteration of factor structure.

Confirmatory Factor Analysis

Typical factor extraction methods use a matrix of Pearson product–moment correlations; however, this method assumes that variables are at least at the interval level. For ordinal data, such as responses on behavior rating scales, it is more appropriate to use polychoric correlations (Joreskog & Sorbom, 1993). Additionally, if item distributions are similar—as was the case with these data, where all items had positively skewed distributions—spurious factors may emerge as a result. To address these problems, we used linear structural equation modeling (LISREL 8) to perform a CFA on polychoric correlations and an asymptotic covariance matrix to test the results of the PAF solution and the theoretical two-dimensional structure of *DSM–IV* criteria.

Two models were computed: a one-factor model where all items were constrained to load on a single factor, and a two-factor model where odd-numbered items were constrained to load only on a presumed inattention factor and even-numbered items were constrained to load only on a presumed hyperactivity–impulsivity factor. The fit of both models (i.e., the extent to which the models accurately reproduced the observed correlation matrices) was analyzed by an examination of multiple goodness-of-fit indices. The assumption that a two-factor solution (vs. a one-factor solution) significantly improves model fit was tested by (a) computing chi-square estimates of improvement of the one- and two-factor solutions over the null model and (b) comparing goodness-of-fit estimates (for a discussion of goodness-of-fit estimates used in this study along with computational formulas, see Medsker, Williams, & Holahan, 1994).

It should be noted that because there is no single, generally accepted index of model fit (Bollen, 1990) and because very large sample sizes can affect fit indices (Marsh, Balla, & McDonald, 1988), the process is more of an interpretation through analysis and comparison of multiple fit indices than a simple rejection of (or failure to reject) the null hypothesis based on a single test statistic. On the basis of this analysis, one can only infer that a particular solution is a possible model that adequately represents the observed data. For the two-factor model, the correlation between factors was .94. The chi squares for both the one- and two-factor models were significant (a significant chi square indicates that the model in question does not fit the data); however, this was expected because a chi square is influenced by the large sample size. In contrast, the root mean square error of approximation (RMSEA; a measure of discrepancy per degree of freedom; Joreskog & Sorbom, 1993) of both models was near the 0.05 value suggested by Browne and Cudeck (1993), indicating close fit. Additionally, the test of RMSEA was 1.00 for both models, also indicating a good fit. The two-factor model resulted in a significant increase in fit over the one-factor model, $\chi^2(1) = 191, p < .01$.

To test the stability of the two-factor model, the sample was randomly split into two groups. A cross-validation index (CVI) was computed using covariance matrices and unweighted least squares (ULS; Joreskog & Sorbom, 1993). The ULS estimation was used because it has no distributional assumptions. The fit of the two samples was tested using one sample as the calibration sample and the other as the validation sample. The closer the CVI value is to 0, the better the fit. The obtained CVI was .24 (90% confidence interval = .20 to .28) suggesting that the same factor structure held for both samples.

Sex, Age, and Ethnic Group Differences

For the purposes of further analyses, ARS scores were computed for inattention (IA, sum of odd-numbered items), hyperactivity-impulsivity (HI, sum of even-numbered items), and total score (sum of all items).

A 2 (Sex) \times 4 (Age) \times 3 (Ethnic Group) multivariate analysis of variance (MANOVA) was conducted using the three ARS scores as the dependent variables. An alpha level of .01 was used, given the number of analyses conducted. To maximize individual cell size, the age factor was blocked into four levels (5- to 7-years-old, 8- to 10-years-old, 11- to 13-years-old, and 14- to 18-years-old). This blocking scheme was supported by age differences found in an initial one-way MANOVA, with 13 levels of age used as the independent variable. Given the relatively small numbers of Asian American and Native American participants, the effects of ethnic group were examined only for Caucasian, African American, and Hispanic children. Furthermore, because data for all three rating scale scores were highly positively skewed, a logarithmic transformation was conducted prior to the MANOVA. Statistically significant results were obtained for the main effects of sex (Wilk's $\lambda = .98$), $F(3, 3, 569) = 30.25$, $p < .0001$; age (Wilk's $\lambda = .99$), $F(9, 8, 686.16) = 4.97$, $p < .001$; and ethnic group (Wilk's $\lambda = .96$), $F(6, 7, 138) = 21.13$, $p < .0001$.

Separate 2 (Sex) \times 4 (Age) \times 3 (Ethnic Group) univariate analyses of variance (ANOVAs) were conducted for each of the ARS scores, using an alpha level of .01 to control for experimentwise error. Significant interaction effects for Age \times Ethnic Group were found for HI, $F(6, 3, 649) = 3.55$, $p < .01$, and total scores, $F(6, 3, 649) = 3.44$, $p < .01$. Simple effects tests of ethnic group at each level of age followed by Tukey honestly significant difference (HSD) post hoc comparisons were then conducted. Significant ethnic group effects were found for both ARS scores within all four age groups (p 's $< .001$). In the 5- to 7-year-old group, African Americans received higher HI (effect size [ES] $= .40$) and total score (ES = .37) ratings than did Caucasians, with no other between-group differences. This pattern of ethnic group differences also was found within the 8- to 10-year-old (HI ES = .42, total score ES = .41) and 11- to 13-year-old (HI ES = .56, total score ES = .61) age groups. Furthermore, African Americans received higher ratings than did Hispanic participants in these latter two age ranges (HI ES = .58 and total score ES = .57 for 8- to 10-year-olds and HI ES = .52 and total score ES = .57 for 11- to 13-year-olds). Alternatively, both African American and Hispanic adolescents in the 14- to 18-year-old group received higher ADHD ratings than did Caucasians (HI ES = .44 and .67 for African Americans and Hispanics, respectively; total score ES = .50 and .65 for African Americans and Hispanics, respectively). To further illustrate these findings, the pattern of ethnic group differences in total score across age groups is portrayed in Figure 1 .

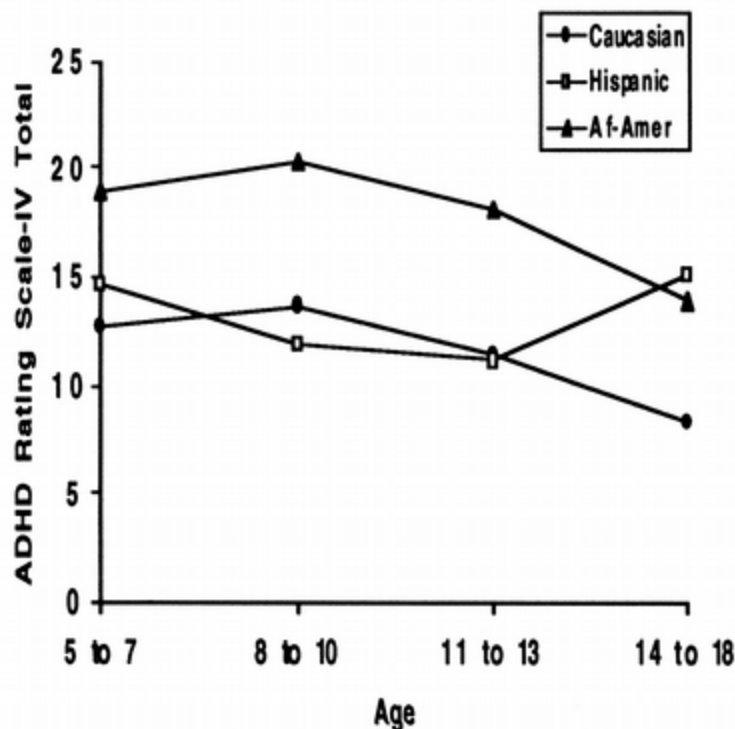


Figure 1. Mean ADHD Rating Scale–IV (School Version) total score across four age groups as a function of ethnic group. Af Amer = African American. ADHD = attention deficit hyperactivity disorder

Significant main effects for sex were found for IA, $F(1, 3, 571) = 76.12, p < .0001$; HI, $F(1, 3, 571) = 78.80, p < .0001$; and total scores, $F(1, 3, 571) = 87.48, p < .0001$. Boys received higher ratings of ADHD symptoms than did girls for HI (ES = .50), IA (ES = .46), and total scores (ES = .52). A significant main effect for age was obtained for HI scores only $F(3, 3, 571) = 7.54, p < .0001$. Tukey HSD post hoc comparisons revealed that 14- to 18-year-olds received significantly lower ratings of HI symptoms than did the three younger age groups (ES ranged from .31 to .54). In similar fashion, 11- to 13-year-olds received lower HI ratings than did children in the 5- to 7- and 8- to 10-year-old age groups (ES = .23 and .19, respectively). Finally, significant main effects for ethnic group were found for IA, $F(2, 3, 571) = 49.27, p < .0001$; HI, $F(2, 3, 571) = 58.44, p < .0001$; and total scores, $F(2, 3, 571) = 59.25, p < .001$. Tukey HSD tests indicated that African American participants received higher ratings on all three ARS scores than did Caucasian (ES ranged from .44 to .48) and Hispanic participants (ES ranged from .31 to .35), with no differences between the latter two groups.

Normative Data

Normative data are provided separately for boys and girls in Tables 2 and 3, respectively. Within each table, means and standard deviations are provided for three scores in accordance with factor analytic results: inattention (IA, sum of nine odd-numbered items), hyperactivity–impulsivity (HI, sum of nine even-numbered items), and total score (sum of inattention and hyperactivity–impulsivity scores). Given the age differences obtained in prior analyses, separate normative data are provided for four age groups (5- to 7-year-olds, 8- to 10-year-olds, 11- to 13-year-olds, and

14- to 18-year-olds). Because very few cases were available for 4- and 19-year-olds, these ages were not included in the normative data set. Separate norms for each ethnic group are not provided because cell sizes (Sex \times Age \times Ethnic Group) would be too small and further research is necessary prior to the presentation of ethnic group norms. Scores are provided for four cutoff points: 80th, 90th, 93rd, and 98th percentile. These cutoff points can be used for screening (80th and 90th percentile) or identification (93rd and 98th percentiles) purposes.

Discussion

Factor analyses of the ARS generally confirmed the two-factor solution of ADHD symptoms proposed in the *DSM-IV* (inattention and hyperactivity-impulsivity) and as conceptualized in contemporary theories of the disorder (Barkley, 1997). The findings demonstrated differences in teacher ratings of ADHD symptoms as a function of sex, age, and ethnic group. In general, the effect sizes associated with each of these variables were in the low to moderate range (i.e., from .19 to .54), with the most pronounced differences accounted for by sex and ethnic group.

Factor Structure of ADHD Ratings

Factor analytic procedures demonstrated that both the one-factor and two-factor models adequately represented the observed data. Although the one-factor model offers the advantage of parsimony, the two-factor model supports the current theoretical conceptualization of ADHD. Moreover, the observed factor structure closely approximates simple structure, so the factors appear quite distinct. In addition, the two-factor solution does explain a statistically significant (although small in the absolute sense) increment in the variance. It also has the advantage of clinical value because it is known on the basis of *DSM-IV* field trial data that children may be high on one dimension (i.e., either inattention or hyperactivity-impulsivity) without being high on the other (Lahey et al., 1994). This is perhaps the most compelling rationale for the two-factor model.

Factor analyses provided additional validation of items selected for the inattention and hyperactivity-impulsivity dimensions of the *DSM-IV*. In each case, items demonstrated a loading greater than or equal to .50 on the factor to which it was hypothesized to belong. Although two items displayed relatively high (i.e., $\geq .30$) loadings on both factors, for each item the loading was higher on the hypothesized factor. More interesting, the three items that purportedly reflect impulse control (i.e., items 14, 16, and 18) had the highest loadings on the HI factor, which is in keeping with the theory that behavioral disinhibition is the core deficit associated with ADHD (Barkley, 1997).

Sex, Age, and Ethnic Group Differences

These findings are consistent with prior research demonstrating sex differences in teacher ratings of inattention and hyperactivity-impulsivity (DuPaul, 1991; Goyette, Conners, & Ulrich, 1978; Trites, Blouin, & Laprade, 1982). In the present study, boys were reported to exhibit significantly more symptoms of ADHD than were girls across all of the age ranges included in the sample. Given the ubiquity of gender differences in the literature, perhaps future versions of the *DSM* should include separate symptomatic thresholds for diagnosing ADHD in girls and boys. Specifically, boys may be required to exhibit a greater number of symptoms than girls to be diagnosed with ADHD.

Our results confirm those of previous studies showing a decline in teacher ratings of hyperactivity–impulsivity with increasing age and the stability of teacher ratings of inattention through childhood and adolescence (Goyette et al., 1978; Hart et al., 1995). Ratings of hyperactivity–impulsivity reflected essentially no change in the age ranges from 5 to 10 years old, roughly coinciding with the elementary school years, declined in the middle school years (ages 11–13), and decreased further in the high school years (14–18). The results suggest that symptoms of hyperactivity–impulsivity generally are less evident in the older groups as compared with the younger groups and that symptoms of inattention are uniformly problematic for elementary, middle school, and high school students. An alternative hypothesis to explain the decline in ratings of hyperactivity–impulsivity is that the *DSM–IV* symptoms reflect overt manifestations of hyperactivity and impulsivity that are more likely to be exhibited by preadolescents than by teenagers. Perhaps teachers of high school-aged students would endorse items that reflect the more covert symptoms of hyperactivity and impulsivity in the adolescent population (e.g., making decisions without forethought). Although our results suggest developmental trends in ADHD symptomatology, firm conclusions are limited by the use of a cross-sectional design.

African American children were rated as, consistent with prior research, significantly more hyperactive than were Caucasian students (Langsdorf, Anderson, Waechter, Madrigal, & Juarez, 1979; Waechter, Anderson, Juarez, Langsdorf, & Madrigal, 1979). In the present study, teachers rated African American students higher than Caucasian students on both the hyperactivity–impulsivity and inattention factors across all age ranges represented in this national sample. Differences between Hispanic and Caucasian students were found only in the 14- to 18-year-old group, with Hispanic students scoring higher. The meaning of these ethnic group differences is unclear and needs to be explored further. In future research, the effects of socioeconomic status, location of residence (i.e., urban vs. suburban vs. rural), and potential item bias need to be examined when comparing ratings among ethnic groups. Also, the ethnic match between the teacher and student may be an important variable to consider.

The normative data reported in this study can be clinically useful in the assessment of children and adolescents for ADHD. As part of a multiple gate assessment process, the ARS could be used as a screening tool to identify students who may meet criteria for ADHD and need a more comprehensive assessment (Power & Ikeda, 1996). In this case, it is recommended that clinicians use a cutoff score that yields very few false negatives (e.g., 90th percentile). The ARS also can be used as a part of a comprehensive diagnostic evaluation of ADHD. Given the limitations of rating scales (Reid & Maag, 1994), it is important to use teacher rating scales in combination with other measures, such as structured diagnostic interviews, direct observations of behavior, and office-based assessment of child functioning as well as parent rating scales (Barkley, 1990; DuPaul & Stoner, 1994; Hinshaw, 1994). When used in diagnostic assessment, selection of a cutoff score that achieves the optimal balance between maximizing true positives while minimizing false negatives is required (Chen, Faraone, Beiderman, & Tsuang, 1994). Future research investigating the predictive power of the ARS using various cutoff scores in community and clinical samples will help to determine the optimal cutoff score for diagnostic assessment. Preliminary results indicate that when the ARS is used as part of a multimethod assessment battery with other types of measures mentioned above, then selection of a cutoff score at the 80th (for predicting inattentive subtype) or 90th (for predicting hyperactive–impulsive subtype)

percentile appears most appropriate (Power et al., 1997). When used in research, if it is critical to ensure that participants meet criteria for ADHD, a highly conservative approach using the 98th percentile as a cutoff score may be indicated. Given the high score needed at some age groups to meet this cutoff (e.g., scores for 5–7-year-old boys in Table 2), these normative data may prove useful in situations where the clinician or researcher seeks to avoid overidentification of ADHD.

Table 2
Normative Data for Boys From the ADHD Rating Scale–IV (School Version)

Age (in years)	n	Inattention				Hyperactivity–impulsivity				Total score									
		M	SD	Percentile				M	SD	Percentile				M	SD	Percentile			
				80th	90th	93rd	98th			80th	90th	93rd	98th			80th	90th	93rd	98th
5–7	243	8.75	7.66	16.0	21.0	22.0	26.1	8.12	7.86	16.0	20.0	22.0	27.0	16.87	14.61	30.2	39.0	41.0	51.0
8–10	307	10.33	8.49	19.0	24.0	25.0	27.0	8.43	8.05	16.0	22.2	25.0	27.0	18.76	15.51	34.0	44.2	46.0	52.8
11–13	221	9.33	8.11	17.0	22.8	24.0	27.0	5.96	6.72	12.0	17.0	18.0	24.6	15.28	13.55	28.0	35.8	37.9	49.1
14–18	223	8.25	7.27	15.0	19.6	21.3	26.5	4.37	6.09	8.0	13.0	17.3	21.0	12.62	12.16	23.0	31.0	34.0	44.0

Note. ADHD = attention deficit hyperactivity disorder.

Limitations and Conclusions

Several limitations of this research should be noted. Although the normative sample closely matches census data from 1990, the western part of the United States is somewhat underrepresented and some ethnic groups, in particular Native Americans and Asian Americans, are also slightly underrepresented. Moreover, in collecting data for this national sample, we did not measure directly some potentially important variables including socioeconomic status.

At this point, clinicians should exercise caution when applying normative data in the assessment of children from ethnic minority groups, especially African Americans. Although a similar two-dimensional factor solution for the symptoms of ADHD has now been found in several different countries, the validity of this structure for each of the ethnic groups represented in the United States needs to be demonstrated. Also, the meaning and differential validity of factors derived for each ethnic group needs to be determined (Reid, 1995). Furthermore, normative data need to be collected for each of the validated factors for a particular ethnic group. In validating and standardizing the factors for clinical use, researchers must determine whether variables, such as socioeconomic status, setting of residence, and ethnic match between teacher and student, have an impact on factor scores.

In conclusion, this study supports a two-factor solution of ADHD symptoms and the selection of items for each dimension of the *DSM–IV*. The results confirmed the presence of sex differences for the inattention and hyperactivity–impulsivity dimensions and a decline in teacher ratings of hyperactivity–impulsivity but not inattention as a function of age. Normative data for this measure should prove useful for both clinical and research purposes.

Footnotes

¹ Between-group effect sizes for sex, age, and ethnic group were calculated using a formula provided by Glass, McGaw, and Smith (1981) as follows: M Group 1 (e.g., boys) minus M Group 2 (e.g., girls) divided by the pooled standard deviation. This statistic provides an estimate of the size of mean differences between groups in terms of standard deviation units.

References

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- 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 ADHD. *Psychological Bulletin*, *121*, 65–94.
- Bauermeister, J. J., Bird, H. R., Canino, G., Rubio-Stipec, M., Bravo, M., & Alegria, M. (1995). Dimensions of attention deficit hyperactivity disorder: Findings from teacher and parent reports in a community sample. *Journal of Clinical Child Psychology*, *24*, 264–271.
- Baumgaertel, A., Wolraich, M. L., & Dietrich, M. (1995). Comparison of diagnostic criteria for attention deficit disorders in a German elementary school sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34*, 629–638.
- Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin*, *107*, 256–259.
- Brito, G. N. O., Pinto, R. C. A., & Lins, M. F. C. (1995). A behavioral assessment scale of attention deficit disorder in Brazilian children based on *DSM-III-R* criteria. *Journal of Abnormal Child Psychology*, *23*, 509–521.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structure equation models* (pp. 136–162). London: Sage.
- Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist Scales for attention-deficit hyperactivity disorder: A receiver-operating analysis. *Journal of Consulting and Clinical Psychology*, *62*, 1017–1025.
- Conners, C. K. (1989). *Conners' Rating Scales manual*. North Tonawanda, NY: Multi-Health Systems.
- DuPaul, G. J. (1991). Parent and teacher ratings of ADHD 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, Lehigh University.
- DuPaul, G. J., Power, T. J., McGoey, K., Ikeda, M., & Anastopoulos, A. D. (1997, April). *Reliability and validity of the AD/HD Rating Scale-IV*. Poster session presented at the annual convention of the National Association of School Psychologists, Anaheim, CA.
- DuPaul, G. J., & Stoner, G. (1994). *ADHD in the schools: Assessment and intervention strategies*. New York: Guilford Press.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Gorsuch, R. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
- Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on Revised Conners Parent and Teacher Rating Scales. *Journal of Abnormal Child Psychology*, *6*, 221–236.
- Hart, E. L., Lahey, B. B., Loeber, R., Applegate, B., & Frick, P. J. (1995). Developmental change in attention-deficit hyperactivity disorder in boys: A four-year longitudinal study. *Journal of Abnormal Child Psychology*, *23*, 729–750.

- Hinshaw, S. P. (1994). *Attention deficits and hyperactivity in children*. Thousand Oaks, CA: Sage.
- Joreskog, K., & Sorbom, D. (1993). *LISREL 8*. Hillsdale, NJ: Erlbaum.
- Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., Barkley, R. A., Newcorn, J., Jensen, P., Richters, J., Garfinkel, B., Kerdyk, L., Frick, P. J., Ollendick, T., Perez, D., Hart, E. L., Waldman, I., & Schaffer, D. (1994). DSM-IV field trials for attention-deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*, *151*, 1673–1685.
- Lahey, B. B., & Carlson, C. (1992). Validity of the diagnostic category of attention deficit disorder without hyperactivity: A review of the literature. In S. E. Shaywitz & B. A. Shaywitz (Eds.), *Attention deficit disorder comes of age: Toward the twenty-first century* (pp. 119–144). Austin, TX: Pro-Ed.
- Lahey, B. B., Pelham, W. E., Schaughency, E. A., Atkins, M. S., Murphy, H. A., Hynd, G. W., Russo, M., Hartdagen, S., & Lorys-Vernon, A. (1988). Dimensions and types of attention deficit disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *27*, 330–335.
- Langsdorf, R., Anderson, R., Waechter, D., Madrigal, J. F., & Juarez, L. J. (1979). Ethnicity, social class, and perception of hyperactivity. *Psychology in the Schools*, *16*, 293–298.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410.
- McCarney, S. B. (1989). *The Attention Deficit Disorders Evaluation Scale: School version technical manual*. Columbia, MO: Hawthorne Educational Services.
- Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, *20*, 439–464.
- 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.
- Power, T. J., Doherty, B. J., Panichelli-Mindel, S. M., Karustis, J. L., Eiraldi, R. B., Anastopoulos, A. D., & DuPaul, G. J. (1997). *Integrating parent and teacher reports in the diagnostic assessment of ADHD*. Manuscript submitted for publication.
- Power, T. J., & Ikeda, M. J. (1996). The clinical utility of behavior rating scales: Comments on the diagnostic assessment of AD/HD. *Journal of School Psychology*, *34*, 379–385.
- Reid, R. (1995). Assessment of ADHD with culturally different groups: The use of behavior rating scales. *School Psychology Review*, *24*, 537–560.
- Reid, R., & Maag, J. W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology*, *32*, 339–354.
- Shaywitz, S. E., Shaywitz, B. A., Schnell, C., & Towle, V. R. (1988). Concurrent and predictive validity of the Yale Children's Inventory: An instrument to assess children with attentional deficits and learning disabilities. *Pediatrics*, *81*, 562–571.
- Taylor, E., & Sandberg, S. (1984). Hyperactive behavior in English schoolchildren: A questionnaire survey. *Journal of Abnormal Child Psychology*, *12*, 143–156.
- Trites, R. L., Blouin, A. G., & Laprade, K. (1982). Factor analysis of the Conners Teacher Rating Scale based on a large normative sample. *Journal of Consulting and Clinical Psychology*, *50*, 615–623.
- Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1985). Introduction to the use of the ACTeRS. *Psychopharmacology Bulletin*, *21*, 915–920.

Waechter, D., Anderson, R., Juarez, L. J., Langsdorf, R., & Madrigal, J. F. (1979). Ethnic group, hyperkinesis, and modes of behavior. *Psychology in the Schools, 16*, 435–439.

Weiss, G., & Hechtman, L. (1993). *Hyperactive children grown up: AD/HD in children, adolescents, and adults* (2nd ed.). New York: Guilford Press.

Appendix

Attention Deficit Hyperactivity Disorder Rating Scale–IV—School Version

Child's name _____ Age _____ Grade _____

Completed by: _____

Circle the number that *best describes* this student's school behavior over the past 6 months (or since the beginning of the school year).

	Never or rarely	Sometimes	Often	Very often
1. Fails to give close attention to details or makes careless mistakes in schoolwork.	0	1	2	3
2. Fidgets with hands or feet or squirms in seat.	0	1	2	3
3. Has difficulty sustaining attention in tasks or play activities.	0	1	2	3
4. Leaves seat in classroom or in other situations in which remaining seated is expected.	0	1	2	3
5. Does not seem to listen when spoken to directly.	0	1	2	3
6. Runs about or climbs excessively in situations in which it is inappropriate.	0	1	2	3
7. Does not follow through on instructions and fails to finish work.	0	1	2	3
8. Has difficulty playing or engaging in leisure activities quietly.	0	1	2	3
9. Has difficulty organizing tasks and activities.	0	1	2	3
10. Is "on the go" or acts as if "driven by a motor."	0	1	2	3
11. Avoids tasks (e.g., schoolwork, homework) that require sustained mental effort.	0	1	2	3
12. Talks excessively.	0	1	2	3
13. Loses things necessary for tasks or activities.	0	1	2	3
14. Blurts out answers before questions have been completed.	0	1	2	3
15. Is easily distracted.	0	1	2	3
16. Has difficulty awaiting turn.	0	1	2	3
17. Is forgetful in daily activities.	0	1	2	3
18. Interrupts or intrudes on others.	0	1	2	3

Note. From *AD/HD Rating Scale–IV*, by G. J. DuPaul, A. D. Anastopoulos, T. J. Power, K. Murphy, and R. A. Barkley, 1994, Unpublished rating scale, Lehigh University. Copyright 1994 by DuPaul et al. Reprinted with permission.