

## GENDER AND ETHNIC DIFFERENCES IN ADHD AS ASSESSED BY BEHAVIOR RATINGS

By: ROBERT REID, CYNTHIA A. RICCIO, ROBERT H. KESSLER, GEORGE J. DUPAUL, THOMAS J. POWER, ARTHUR D. ANASTOPOULOS, DIANA ROGERS-ADKINSON, AND MARY-BETH NOLL

Reid, R., Riccio, C.A., Kessler, R.H., DuPaul, G.J., Power, T.J., [Anastopoulos, A.D.](#), Rogers-Adkinson, D., & Noll, M.B. (2000). Gender and ethnic differences in ADHD as assessed by behavior ratings. *Journal of Emotional and Behavioral Disorders*, 8, 38-48.

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

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

### **Abstract:**

Attention-deficit/hyperactivity disorder (ADHD) is a common childhood disorder. Research suggests that ADHD is 4 to 9 times more frequent in males than females, and the possibility of underidentification in females and overidentification in males has been suggested as an explanation for these statistics. As part of the diagnostic process, teachers are frequently asked to complete behavior rating scales. There is a lack of empirical data concerning the extent to which gender differences are evident on such rating scales. This study investigated the use of the ADHD-IV Rating Scale--School Version, with male and female students from ages 5 to 18 years. Results suggest that the ADHD construct is consistent across gender; however, there are differences across gender and ethnicity. For Caucasian children, externalizing behaviors are most salient in terms of discriminating between males and females. Implications for research and practice are discussed.

### **Article:**

Attention-Deficit/Hyperactivity disorder (ADHD) is a common childhood disorder that is frequently brought to the attention of physicians and psychologists to conduct evaluations due to behavioral concerns of both teachers and parents (Breen & Altepeter, 1990). According to the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV; American Psychiatric Association, 1994), ADHD is characterized by a pattern of inattention and/or hyperactivity-impulsivity that is exhibited to an extreme level, such that it is developmentally inappropriate relative to a person's age. It occurs in an estimated 3% to 5% of school-age children.

The most evident difference between males and females with ADHD is the higher rate at which males are diagnosed. Male-to-female ratios range from 4:1 to 9:1, depending upon whether community-based or clinical samples are used (APA, 1994). In fact, the behaviors used to define the symptomatology of ADHD in the DSM-IV were identified from a sample pool composed predominately of males (Frick et al., 1994; Lahey et al., 1994). Several hypotheses have been offered to explain the disproportionate frequency of males with ADHD. These hypothesis tend to rely on biological or child-rearing differences by gender (see, e.g., Barkley, 1989; Befera & Barkley, 1984; Brown, Madan-Swain, & Baldwin, 1991; Eme, 1992; Ernst et al., 1994; Gualtieri

& Hicks, 1985; James & Taylor, 1990; Kashani, Chapel, Ellis, & Shekim, 1979), but the issue has not been resolved.

Extensive research has been conducted on boys with ADHD, but comparable studies of girls with the disorder are infrequent, possibly due to the challenges of recruiting sufficient samples of girls with the disorder (Arnold, 1996). Studies on ADHD specific to gender disparity have often been epidemiological in nature (e.g., Barkley, 1989; Berry, Shaywitz, & Shaywitz, 1989; Faraone, Biederman, Keenan, & Tsuang, 1991; James & Taylor, 1990; McGee, Williams, & Silva, 1987). Studies have found more similarities than differences in girls and boys with ADHD, and a number of studies have found no gender differences in the number, or severity, of ADHD symptoms as a function of gender (Berry et al., 1985; Brown et al., 1991; Horn, Wagner, & Ialongo, 1989; James & Taylor, 1990; Kashani et al., 1979; Silverthorn, Frick, Kuper, & Ott, 1996). However, some gender differences related to ADHD have been identified. Seidman and colleagues (1997) found that, in contrast to research findings of impaired executive function in boys with ADHD, girls with ADHD did not demonstrate significant executive function deficits. Other research has found that girls with ADHD show more severe visual-spatial, cognitive, and language deficits than boys with ADHD (Berry et al., 1985; Breen, 1989; Brown et al., 1991; Gordon & Mettelman, 1994; James & Taylor, 1990; Kashani et al., 1979; Taylor, 1986).

The available research on gender differences has revealed some differences with regard to the referral process for boys and girls. Kashani and colleagues (1979) found that females with ADHD were usually referred for learning problems rather than behavior problems, whereas boys identified as hyperactive were more frequently referred for behavior problems than learning problems. This would be consistent with the research findings that girls with ADHD tend to have more comorbid internal manifestations of the disorder, whereas boys with ADHD have been noted to express more aggressive overt types of behavior (Breen & Altepeter, 1990; Brown, Abramowitz, Madan-Swain, Eckstrand, & Dulcan, 1989; deHaas, 1986; Gordon & Mettelman, 1994). Thus, gender-correlated behavioral patterns may be more frequently identified as ADHD in boys than in girls due to the frequency of disruptive classroom behavior exhibited by boys (Breen & Altepeter, 1990).

One important caution to keep in mind is that gender differences may vary as a function of the sample used. For example, Gaub and Carlson (1997) found a trend for greater severity of inattention among females, but comparable levels of hyperactivity in females and males, in a clinically referred sample. However, this pattern was not evident among nonreferred children.

It has been suggested that one cause for the gender disparity in referrals for ADHD may rest in the scales used as part of the diagnostic process, and the use of general norms as opposed to gender-specific norms (Barkley, 1996). Gender differences have been noted on rating scales used to assess children suspected of having ADHD. Based on information provided in the manual for the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992), for the standardization sample, males were rated one third to one half of a standard deviation higher than females on hyperactivity, attention problems, and other subscales. For this reason, the BASC manual includes separate norm tables for males and females, as well as a combined norm table, and recommends that the same-sex norms be used for clinical diagnosis in order to identify those children whose ratings are significant for both their age and gender. Similarly, DuPaul and

colleagues (1997) reported significant differences across gender on composite inattention and hyperactivity scores based on normative data from the ADHD-IV Rating Scale-School Version (ARS; DuPaul, Power, Anastopoulos, & Reid, 1998). DuPaul et al. (1998) provided separate norms for males and females.

In contrast, Silverthorn and colleagues (1996) concluded that separate norms by gender were not warranted based on their finding that girls and boys with ADHD did not differ on measures of severity and that diagnostic cut-scores identified boys and girls with equivalent levels of impairment. They further argued that to use separate norms might artificially reduce the difference in prevalence rates for ADHD for girls and boys. Gordon (1996) argued that boys are more often identified with ADHD because they are more likely to demonstrate the severity of symptomatology to warrant this diagnosis; therefore, gender referencing is not appropriate. Alternatively, the possibility has been suggested that ADHD manifests differently in girls as compared to boys (Barkley, 1996; Gordon & Mettelman, 1994).

When considering the possibility of differences in ADHD symptomatology across gender, ethnicity is a factor that should also be considered. Serious concerns have been expressed regarding the assessment of ADHD with children from ethnic minorities (Bauermeister, Berrios, Jimenez, Acevedo, & Gordon, 1990) and the use of behavior rating scales with these children (Reid, 1995). However, ADHD among ethnic minorities remains an understudied area. Studies that have used behavior rating scales have reported significant differences between Caucasian (see Note) and African American students (e.g., DuPaul et al., 1997; Epstein, March, Conners, & Jackson, 1998; Jarvinen & Sprague, 1995; Lambert, Sandoval, & Sassone, 1978; Reid et al., 1998; Waechter, Anderson, Juarez, Langsdorf, & Madrigal, 1979). However, it is uncertain whether these differences were due to real differences in behavior among groups, rarer bias due to ethnicity and/or socioeconomic factors, or a combination of the two. Only one study to date has addressed the issue of rarer bias. This study found evidence that ethnic status appeared to bias behavior ratings (Sonuga-Barke, Minocha, Taylor, & Sandberg, 1993). Because ethnic minorities constitute a large portion of the population, and because research has demonstrated convincingly that there are interethnic differences on behavior rating scales, we should attend to ethnicity as a factor in gender differences.

The purpose of this study was to explore the extent and nature of gender differences in ADHD across different ethnic groups (i.e., Caucasian and African American). Although clinical diagnosis requires multimethod, multisource information, we focused on teacher-completed rating scale information, because teacher ratings are often viewed as indicative of functioning in the school environment and are, thus, an important component in the assessment process (e.g. Barkley, 1989). Our exploration involved four interrelated analyses:

1. We analyzed descriptive statistics.
2. We performed a multivariate analysis to assess whether or not different patterns of item means existed across gender for two ethnic groups.
3. We used structural equation modeling to investigate whether the same two-factor construct of ADHD was appropriate for both males and females.
4. We conducted a discriminant function analysis to determine the extent to which items representing ADHD symptoms best separated children by gender.

Because of the possibility that gender differences may vary depending on symptom severity, we conducted analyses on two samples: one representative of unselected children and a second representing children with severe or numerous symptoms of ADHD.

## METHOD

### *Instrumentation*

Teachers completed the ARS. The ARS was selected because it reflects current DSM-IV diagnostic criteria for ADHD. This scale consists of 18 items--9 that address hyperactivity-impulsivity (HI) symptoms and 9 that address inattention (IA) symptoms--directly adapted from the diagnostic criteria for ADHD specified in the DSM-IV. The ARS has demonstrated adequate reliability and criterion-related validity (DuPaul, Power, McGooley, Ikeda, & Anastopoulos, 1998).

In order to minimize possible bias due to response set, IA symptoms were designated as odd-numbered items, and HI symptoms were designated as even-numbered items. Teachers were instructed to select 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 for each item was delineated on a four-point Likert scale ranging from 0 (never or rarely) to 3 (very often), with the higher scores indicative of greater ADHD-related behavior. Teachers were asked to rate the behavior of two students from their class roster (e.g., third male and fifth female on the class roster). Ratings were completed between October and May in the 1994-1995 or 1995-1996 school years. Estimated return rates ranged from 50% to 95% ( $M = 85\%$ ) across school districts.

### *Participants*

Participants for this study included 3,322 children and youth ages 5 to 18 (referred to as the Total group) taken from the normative sample of the ARS (DuPaul et al., 1997). The Caucasian (CA) sample consisted of 1,338 males and 1,298 females. The African American (AA) sample consisted of 376 males and 310 females. Because the total behavior rating scores decrease as age levels increase, the possibility of an age by gender dependency was tested for both CA and AA ethnic groups. The results suggested that ages were proportional across gender for both groups, AA group,  $\chi^2(12, N = 686) = 7.14, p = .84$ ; CA group,  $\chi^2(12, N = 2636) = 12.39, p = .71$ . A subsample of participants who might have been considered at risk for ADHD (referred to as the At-Risk group)--who scored above the 90th percentile on HI, IA, or both factors--was created. The 90th percentile was selected because it is the recommended cut-score (DuPaul, Power, Anastopoulos, & Reid, 1998). This resulted in an At-Risk group consisting of 218 CA males, 203 CA females, 64 AA males, and 53 AA females.

Participating teachers were largely women (82.0%) and Caucasian (93.4%). Other ethnic groups represented among teachers included African American (5.4%), Hispanic (0.7%), Native American (0.1%), Asian American (0.1%), and other (0.4%). The majority of teachers (91.1%) were general educators; the remainder (8.9%) were special educators. No dependency was found between special education or general education status and gender for either ethnic group, AA group,  $\chi^2(1, N = 686) = .43, p = .51$ ; CA group,  $\chi^2(1, N = 2636) = .04, p = .85$ .

## *Analysis*

**Descriptive Data.** Descriptive data analysis followed the guidelines suggested by Bracken and Barona (1991). Descriptive statistics for each group were computed along with reliability coefficients (Cronbach's alpha) for IA and HI factors. To assess the magnitude of difference across gender, effect sizes for each item were computed for both the CA and AA groups. Because effect sizes are a standardized measure, they allow for direct comparison of differences. Effect sizes were calculated by subtracting the female mean from the male mean, then dividing by the square root of the pooled variance.

**Multivariate Analysis.** To test for differences in item means across gender and ethnicity, a 2 (gender) x 2 (ethnicity) multivariate analysis of variance was performed. Separate analyses were conducted for the Total group and the At-Risk group.

**Structural Equation Modeling.** This analysis was conducted only for the Total group, because the number of AA males and females in the At-Risk group was too small. To test whether the two-factor (HI and IA) model of ADHD based on the DSM-IV was consistent across gender, structural equation modeling (SEM) was used to compare the factor structure, factor correlations, item loadings, and item uniqueness across gender, following the procedures (LISREL 8) suggested by Joreskog and Sorbom (1993) and Benson (1987).

The two-factor model was shown to adequately model observed data when factor analysis was conducted on the ARS during the norming process (DuPaul, Power, Anastopoulos, & Reid, 1998). To determine if the factor structure was invariant across genders, a four-step procedure was used. First, males and females were compared with all parameters (i.e., item factor loadings, factor correlations, and item uniqueness) constrained to be equal to those of the males. This step provides a baseline estimate of model fit. Second, separate estimates of item-factor loadings were computed for males and females. Third, separate estimates of item-factor loadings and factor correlations were computed for males and females. The last step involved computing the separate estimates of factor loadings, factor correlations, and item uniqueness (random measurement error and unique item variance) for males and females. The results from Steps 2 through 4 were then compared to the results of Step 1 to assess model fit with each additional freed parameter. The chi-square difference test was used to determine whether freeing a restraint improved the overall fit of the model (Bentler & Bonett, 1980). All parameter estimates were performed using covariance matrices and generalized least-square estimation.

**Discriminant Function Analysis.** To determine the extent to which ARS items were predictive of gender, we conducted discriminant function analysis using a procedure suggested by Huberty (1984). A step-wise analysis was used to determine the most parsimonious group of predictor variables and to analyze each variable in terms of its effect on between-gender discrimination. Separate analyses were conducted for the Total group and the At-Risk group.

## RESULTS

### *Descriptive Data*

Item means for the Total group and the At-Risk group are shown in Figures 1 and 2. The reliability of the scale for all groups was consistently high. Cronbach's alpha for both the HI and IA scales was greater than .90 for AA and CA participants for both the Total and At-Risk groups.

Table 1 shows between-gender effect sizes by item, separately for CA and AA ethnic groups within the Total group and the At-Risk group. In general, between-gender differences were slightly greater for the CA group than for the AA group. However, for some items such as Item 7 (fails to finish work), this was not the case.

For most items, between-gender differences tended to be higher for the At-Risk group than for the Total group. However, in the case of the AA At-Risk group, between-gender differences decreased markedly for four of the nine IA items: 1 (fails to attend/careless), 5 (doesn't listen), 7 (fails to finish work), and 9 (difficulty with organization). As Table 1 shows, the average between-gender effect sizes for the Total group were CA = .40, AA = .36, and for the At-Risk group, CA = .52, AA = .34.

### *Multivariate Analysis*

For the Total group, significant effects were found both for gender,  $F(18,3,310) = 12.66$ ,  $p < .001$ ;  $\eta^2 = .065$ , and ethnicity,  $F(18,3,310) = 9.75$ ,  $p < .001$ ;  $\eta^2 = .053$ . All univariate tests were significant ( $p < .001$  for all items) for both gender and ethnicity. A significant gender by ethnicity interaction was also found,  $F(18,3,301) = 2.57$ ,  $p < .001$ ;  $\eta^2 = .014$ . However, none of the univariate tests attained significance. Only two items approximated the .05 level (Item 4 [leaves seat],  $p = .064$ , and Item 7 [fails to finish work],  $p = .085$ ). For the At-Risk group, significant effects also were found both for gender,  $F(18,511) = 5.70$ ,  $p < .001$ ;  $\eta^2 = .167$ , and ethnicity  $F(18,511) = 5.82$ ,  $p < .001$ ;  $\eta^2 = .170$ . With the exception of Item 7 (fails to finish work), all univariate tests were significant ( $p < .05$ ) for both gender and ethnicity. The gender by ethnicity interaction was not significant,  $F(18,511) = 1.36$ ,  $p = .15$ .

Because the MANOVA procedure uses a weighted composite to assess interactions, analysis is not straightforward. The interaction for the total sample appears to be due to differences among a small number of items, in combination with a wider separation of group centroids, between the CA groups than the AA groups. However, the practical significance of the interaction is questionable. Power was high, and the interaction effect sizes ( $\eta^2$ ) were in the small range (Stevens, 1996).

### *SEM Analysis*

Tables 2 and 3 show the results of the SEM analysis. The results at Step one, where all parameters are invariant, show that all fit measures for both the CA (Table 2) and AA (Table 3) groups are at levels suggesting an acceptable fit across gender. The goodness of fit index is near the .90 level for both ethnic groups, suggesting an adequate fit. Also, the comparative fit index and the non-normed fit index are well above .90 for both groups. Additionally, the root mean square error of approximation is at or below the .08 level suggested as indicative of acceptable fit (Joreskog & Sorbom, 1993). This implies that, from a measurement perspective, there are no major qualitative distinctions across gender for either the CA or the AA group. When separate estimates of item uniquenesses were made, there was a statistically significant decrease in chi-square for both groups. However, because the fit at Step one, where all parameters were constrained across gender, is acceptable, the most conservative interpretation is that the same two-factor ADHD model holds across gender for both ethnic groups.

### *Discriminant Function Analysis*

**CA Group.** For the Total group, a significant discriminant function for gender with eight predictors was obtained,  $\chi^2(8, N = 2636) = 302.36, p < .0001$ . Univariate statistics for each predictor are presented in Table 4. Group centroids for the male and female groups were .344 and -.354, respectively. For the At-Risk group, a significant discriminant function for gender with five predictors was also obtained,  $\chi^2(5, N = 421) = 121.86, p < .0001$ . Univariate statistics for each predictor are presented in Table 4. Group centroids for the male and female groups were .561 and -.612, respectively.

**AA Group.** For the Total group, a significant discriminant function for gender with four predictors was obtained,  $\chi^2(4, N = 686) = 48.09, p < .0001$ . Univariate statistics for each predictor are presented in Table 5. Group centroids for the male and female groups were .245 and -.297, respectively. For the At-Risk group, a significant discriminant function with two predictors was also obtained,  $\chi^2(2, N = 117) = 19.67, p < .0001$ . Univariate statistics for each predictor are presented in Table 5. Group centroids for the male and female groups were .396 and -.471, respectively.

## DISCUSSION

The results of this study suggest that gender has a significant effect on teacher ratings of ADHD symptomatology. Results also suggest that there are no significant qualitative differences in symptomatology across gender and that the ADHD construct is consistent across gender. However, the results also suggest the possibility of cross-gender differences based on ethnicity.

### *Uniqueness of Symptomatology*

If ADHD symptoms manifest differently across gender, then we would expect to see either a different pattern of item means and effect sizes, or differences in item variance/covariance structure. The results of this study suggest just the opposite. The observed pattern of item means was strikingly similar across gender and ethnicity. As Figures 1 and 2 show, there is a consistent pattern: the AA males were seen as most severe, the AA females and CA males were indistinguishable and distinctly separate from the AA males, and the CA females were seen as least severe and were distinctly separate from the AA females and CA males. The between-gender item effect sizes, which incorporate item variance for each group, further indicate that differences were consistent across items.

If females at risk for ADHD manifest higher rates of inattentive or hyperactive behavior, we would expect to see smaller between-gender item effect sizes for the At-Risk group than for the Total group for items on the IA or HI factors. However, this result was not generally obtained. Aside from smaller effect sizes for a few items for the AA At-Risk group (see Table 1), there appear to be no consistent pattern of item effect size differences between the Total group and the At-Risk group across either HI or IA factors.

Finally, the results of the SEM analyses suggest that, in terms of factor structure, factor loadings, and factor correlations, there is no difference between gender across ethnic groups. Thus, the results of both the descriptive analyses and the SEM analyses strongly suggest that, within the CA and AA groups, the ADHD construct was consistent across gender for the total sample. We should caution that, because the number of AA students in the male and female At-Risk groups fell below the limits of a minimum of 100 participants per group (Loehlin, 1992), we did

not use SEM to compare the At-Risk groups' structures. Thus, the results of the SEM analyses are limited to the Total groups. It is possible that structural differences exist between males and females in the more severely involved groups.

### *Effects of Gender on Teacher Ratings*

The results of this study are consistent with previous analyses of the effects of gender on behavioral ratings (e.g., Reynolds & Kamphaus, 1992; Trites, Blouin, & Laprade, 1982) and strongly support the notion of gender differences in the perceived severity of symptom expression as assessed by teachers' behavior ratings. The MANOVA results indicate significant cross-gender differences when one considers both the Total group and the At-Risk group. The observed between-gender effect sizes for items in Table 1 indicate that these differences are in the moderate range and exist for most scale items. This suggests that perceived gender differences are broad in nature, as opposed to being limited to a subset of items or a single dimension (i.e., IA or HI).

Furthermore, between-gender differences increased slightly in the At-Risk group; thus, gender differences increased as symptom severity increased. This is consistent with the fact that males are more likely to be identified as having ADHD than females. Interestingly, in the MANOVA analyses, when both gender and ethnicity are considered simultaneously, gender accounted for somewhat more variance than ethnicity, as evidenced by the  $\eta^2$  statistics. Thus, the effect of gender appears to be equal to or slightly greater than that of ethnicity on teacher ratings.

In the discriminant function analyses, the items that most distinguished males from females in terms of a unique effect (i.e., when effects of correlation with other variables are parceled out) were not consistent across ethnic groups. For the CA group, a distinct pattern emerged. Items from the HI factor that reflect externalizing behaviors (i.e., fidgets, talks excessively, blurts out answers, difficulty playing quietly, and difficulty organizing tasks) constituted the majority of items and had the greatest effect on male-female separation for both the Total group and the At-Risk group. Thus, for the CA group, externalizing behavior seems to play a major role in defining gender differences. This is likely due, at least in part, to the fact that these behaviors are disruptive to the classroom environment and are more likely to be salient to the classroom teacher. For the AA group, no such pattern emerged. Fewer items attained significance, and only one item was significant for both the Total group and the At-Risk group (loses things necessary for tasks and activities).

This difference between the CA and AA groups appears to be the result of two factors. First, there is less difference between genders for the AA group, as evidenced by the between-gender effect sizes. Second, the shared variance for items appears to be greater for the AA group; thus, collinearity (i.e., high item intercorrelations) becomes a problem. This suggests that teachers' perceptions of the ADHD-related behaviors of AA students is consistent across gender, with few distinguishing behaviors.

Why have some studies that have used behavior ratings (Breen, 1989; Breen & Altpeter, 1990; Horn et al., 1989; James & Taylor, 1990; Silverthorn et al., 1996) failed to find statistically significant between-gender differences? This is likely due to two reasons. First, sample sizes in previous studies were small, ranging from 39 (Breen, 1989) to 80 (Silverthorn et al., 1996).

Thus, some studies may have lacked the statistical power needed to find gender differences of moderate size, such as those identified in the present study. A related problem lies in the small numbers of females with ADHD in the studies noted above, ranging from 13 (Breen, 1989) to 18 (James & Taylor, 1990). Consequently, the stability of results may be an issue. A second reason lies in the participant selection process. The present study used a randomized selection process from a school-based population. In contrast, in studies that failed to find differences, the participants were drawn from clinically referred groups. One study used students identified by ICD-9 standards (James & Taylor, 1990), and another used students with pervasive ADHD (Horn et al., 1989). The reliance on clinically referred groups has been questioned on the grounds that children who were clinically referred may not be representative, as they may constitute an extreme group (Epstein, Shaywitz, Shaywitz, & Woolston, 1991). It is also possible that there is a higher behavioral threshold for females (Eme, 1992). If so, females diagnosed as having ADHD would need to demonstrate extreme levels of behavior to receive an ADHD diagnosis. Therefore, clinically referred females with ADHD in these studies may not be representative of the general population of females with ADHD. However, we cannot rule out the possibility that there may be a subgroup of females who do not differ significantly from males.

### *Same or Different Norms*

Present results have implications for the issue of separate ADHD norms for males and females and/or different ethnic groups. Generally stated, should we make judgments of deviance or disorder on the basis of an absolute standard or a relative standard? An absolute standard implies that a single threshold should be used to define behavior as disordered and that this threshold is valid across potentially mitigating factors, such as gender or ethnicity. A relative standard implies that there should be different thresholds for different groups (e.g., males and females or AA and CA groups) and that deviance or disorder should be judged within the context of group membership.

Results of the present study argue for the relative approach (separate standards for different groups) for using behavior rating scales. The use of an absolute standard implies the use of a single cut-score for all groups. This, in turn, would require pooling the scores of groups that are demonstrably different in terms of means and variances for all scale items. The result of this process is a "hybrid," composed of two dissimilar groups, with a cut-score that would not accurately represent either group. Males would be more likely to screen positive (exceed the cut-score), due to the inclusion of females; females would be less likely to screen positive, due to the inclusion of males. The fact that gender differences (i.e., between-gender effect sizes) were more pronounced as severity increased would further exacerbate this situation.

The problem of accurate representation is compounded when ethnicity is included. Our results demonstrated significant differences in item means across the AA and CA groups (males and females combined). Similarly, mean differences across AA and CA groups have been demonstrated in a number of previous studies and across different behavior rating scales (e.g., DuPaul et al., 1997; Epstein et al., 1998; Jarvinen & Sprague, 1995; Lambert et al., 1978; Reid et al., 1998; Waechter et al., 1979). In addition, for the Total group, we found ethnicity by gender interaction, which apparently was due to two considerations. First, as evidenced by the effect size differences, there was greater separation between genders for the CA group than for the AA group. Second, the discriminant analysis showed that the CA and AA groups differed in terms of

those items that contributed most to gender group separation (see Tables 4 and 5). We note two cautions with regard to ethnicity by gender interaction: The interaction effects were small, and the absence of an interaction for the At-Risk groups suggests that the clinical significance of the interaction is minimal.

Another point in support of separate norms relates to the nature of behavior rating scales. As Barkley (1987) noted, behavior rating scales are "simply quantifications of adult opinions. ..." (p. 219). As a result they are subject to the same sources of unreliability as those opinions. ... A number of potential sources of rater-based error have been identified (Reid & Maag, 1994). One of them--halo effects--has direct relevance to the use of separate norms. Abikoff, Courtney, Pelham, and Koplewicz (1993) reported that halo effects, which inflate behavior rating scale scores, can occur when teachers rate students with oppositional behaviors. Similar findings have been reported by Schachar, Sandberg, and Rutter (1986). Since males are more likely to demonstrate these types of behaviors, they would be more likely to be subject to halo effects and, thus, have spuriously higher ratings.

Halo effects may also differ across ethnic groups. Epstein et al. (1998) reported that items on the Conners Teacher Rating Scale reflecting oppositional behavior (e.g., quarrelsome, defiant, uncooperative) loaded on different factors for AA and CA males. Other studies have also suggested the possibility of halo effects for AA students (Reid et al., 1998) and rater effects for students who are ethnically different (Sonuga-Barke et al., 1993). We see the combination of differences across gender and ethnic groups and the possibility of systematic rater influence on scores as a strong rationale for the use of separate norms for behavioral ratings based on gender, at least, and, perhaps, ethnicity. Rater effects and halo effects are topics requiring additional research.

In summary, the present results show that, although the two-factor ADHD construct is consistent across genders, there are broad-based significant differences in item scores across gender and ethnicity. However, these differences do not appear to be consistent across ethnic groups. For the CA group, a consistent group of items best discriminated males and females in both the total sample and the clinical group; however, no such pattern was evident for the AA group. The results of this study also suggest that separate norms are appropriate for males and females.

We wish to emphasize several cautionary limitations. Present results and arguments are based on teacher ratings on the ARS rating scale. Parent ratings, for example, may produce different results. In addition, differences between ethnic groups could be due to socioeconomic status. Future research should control for this variable. Finally, the value of cut-scores should be determined by the relationships between cut-scores and functional impairment. Differences cannot be used synonymously with deviance. Further research will be required to determine the relationships among gender, ethnicity, socioeconomic status, and functional impairment, as assessed by behavior rating scales.

#### **Note**

For purposes of this study, Caucasian is defined as persons of European-American ancestry.

TABLE 1  
Between-Gender Effect Sizes for 18 Items of the ARS

Item	Total group		At-risk group	
	CA	AA	CA	AA
1. Fails to attend/careless	0.47	0.42	0.52	0.23
2. Fidgets	0.55	0.41	0.65	0.49
3. Difficulty sustaining attention	0.45	0.42	0.62	0.51
4. Leaves seat	0.31	0.39	0.45	0.43
5. Doesn't listen	0.34	0.32	0.42	0.08
6. Runs/climbs excessively	0.36	0.36	0.57	0.62
7. Fails to finish work	0.36	0.48	0.32	0.03
8. Difficulty playing quietly	0.44	0.38	0.72	0.40
9. Difficulty with organization	0.45	0.35	0.66	0.17
10. "On the go"/"Driven"	0.40	0.37	0.67	0.27
11. Avoids sustained mental effort	0.43	0.42	0.46	0.44
12. Talks excessively	0.23	0.23	0.27	0.29
13. Loses things	0.38	0.41	0.55	0.53
14. Blurts answers	0.43	0.29	0.41	0.39
15. Easily distracted	0.44	0.33	0.64	0.35
16. Difficulty awaiting turn	0.43	0.36	0.47	0.37
17. Forgetful	0.36	0.29	0.48	0.08
18. Interrupts/intrudes	0.41	0.35	0.52	0.36
Mean effect size	0.40	0.36	0.52	0.34

Note. ARS = ADHD-IV Rating Scale-School Version (DuPaul, Power, Anastopoulous, & Reid (1998).

TABLE 2  
 Test of Model Fit Across CA Males and Females  
 Legend for Chart:

- A - Model
- B - df
- C -  $\chi^2$
- D -  $df_{(dif)}$
- E -  $\chi^2_{(dif)}$
- F - GFI
- G - RMSEA
- H - CFI
- I - NNFI

A	B	C	D	E
	F	G	H	I
Invariant	305	3111	--	--
	.89	.083	.99	.99
+ Loadings	287	2986	18	25
	.89	.084	.99	.99
+ Factor correlations	286	2985	1	1
	.89	.084	.99	.99
+ Uniqueness	268	2368	18	617[1]
	.91	.076	.99	.99

**Note. GFI = goodness of fit index; RMSEA = root mean square error of approximation; CFI = comparative fit index; NNFI = non-normed fit index.**

*1 p < .01.*

TABLE 3  
 Test of Model Fit Across AA Males and Females  
 Legend for Chart:

- A - Model
- B - df
- C -  $\chi^2$
- D -  $df_{(dif)}$
- E -  $\chi^2_{(dif)}$

F - GFI  
 G - RMSEA  
 H - CFI  
 I - NNFI

A	B	C	D	E
	F	G	H	I
Invariant	305 .84	942 .078	-- .99	-- .99
+ Loadings	287 .85	916 .080	18 .99	26 .99
+ Factor correlations	286 .85	915 .080	1 .99	1
+ Uniqueness	268 .86	832 .078	18 .99	83[1] .99

**Note. GFI = goodness of fit index; RMSEA = root mean square error of approximation; CFI = comparative fit index; NNFI = non-normed fit index.**

*1 p < .01.*

TABLE 4  
 Stepwise Discriminant Function Results for CA Group  
 Legend for Chart:

A - Total group, Item  
 B - Total group, Unique F  
 C - Total group, Standardized discriminant function  
 D - At-risk group, Item #  
 E - At-risk group, Unique F  
 F - At-risk group, Standardized discriminant function

A	B	C	D	E	F
2	50.02	.674	2	5.35	.249
12	48.12	-.654	14	6.60	.300
14	30.16	.489	8	14.03	.404
8	20.34	.411	9	30.14	.568
4	17.97	-.402	15	7.16	.281
1	11.53	.320	--	--	--
11	9.59	.323	--	--	--
7	4.95	-.231	--	--	--

Note. Item 2 = fidget; Item 12 = talks excessively; Item 14 = blurts answers; Item 8 = difficulty playing quietly; Item 4 = leaves seat; Item 1= fails to attend/careless; Item 11 = avoids sustained mental effort; Item 7 = fails to finish work; Item 9 = difficulty with organization; Item 15 = easily distracted.

TABLE 5  
Stepwise Discriminant Function Results for AA Group  
Legend for Chart:

- A - Total group, Item
- B - Total group, Unique F
- C - Total group, Standardized discriminant function
- D - At-risk group, Item #
- E - At-risk group, Unique F
- F - At-risk group, Standardized discriminant function

A	B	C	D	E	F
13	4.17	.495	13	9.60	.700
17	7.69	-.733	6	10.41	.727
2	6.07	.471	--	--	--
7	10.72	.770	--	--	--

Note. Item 13 =loses things;Item 17 = forgetful;Item 2 = fidgets;Item 7 = fails to finish work;Item 13 =loses things;Item 6 = runs/climbs excessively.

GRAPH: FIGURE 1. Item means for the Total group.

GRAPH: FIGURE 2. Item means for the At-Risk group.

### References

- Abikoff, H., Courtney, M., Pelham, W., & Koplewicz, H. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*, 21, 519-533.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Arnold, L. E. (1996). Sex differences in ADHD: Conference summary. *Journal of Abnormal Child Psychology*, 24, 555-569.
- Barkley, R. A. (1987). The assessment of attention deficit-hyperactivity disorder. *Behavioral Assessment*, 9, 207-233.
- Barkley, R. A. (1989). Hyperactive girls and boys: Stimulant drug effects on motherchild interactions. *Journal of Child Psychology and Psychiatry*, 30, 379-390.
- Barkley, R. A. (1996). Gender is already implicit in the diagnosis of ADHD: Shouldn't it be explicit? *ADHD Report*, 4(2), 3-6.
- Bauermeister, J. J., Berrios, V., Jimenez, A. L., Acevedo, L., & Gordon, M. (1990). Some issues and instruments for the assessment of attention-deficit hyperactivity disorder in Puerto Rican children. *Journal of Clinical Child Psychology*, 19, 9-16.

- Befera, M., & Barkley, R. A. (1984). Hyperactive and normal girls and boys: Motherchild interactions, parent psychiatric status, and child psychopathology. *Journal of Child Psychology and Psychiatry*, 26, 439-452.
- Benson, J. (1987). Detecting item bias in affective scales. *Educational and Psychological Measurement*, 47, 55-67.
- Bentler, P.M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Berry, C. A., Shaywitz, S. E., & Shaywitz, B. A. (1985). Girls with attention deficit disorder: A silent minority? A report on behavioral and cognitive characteristics. *Pediatrics*, 76, 801-809.
- Bracken, B., & Barona, A. (1991). State of the art procedures for translating, validating, and using psychoeducational tests in cross-cultural assessment. *School Psychology International*, 12, 119-132.
- Breen, M. J. (1989). Cognitive and behavioral differences in ADHD boys and girls. *Journal of Child Psychology and Psychiatry*, 30, 711-716.
- Breen, M. J., & Altepeter, T. S. (1990). Situational variability in boys and girls identified as ADHD. *Journal of Clinical Psychology*, 46(4), 486-490.
- Brown, R. T., Abramowitz, A.J., Madan-Swain, A., Eckstrand, D., & Dulcan, M. (1989, October). ADHD gender differences in a clinic-referred sample. Paper presented at the meeting of the American Academy of Child and Adolescent Psychiatry, New York.
- Brown, R. T., Madan-Swain, A., & Baldwin, K. (1991). Gender differences in a clinic-referred sample of attention-deficit disordered children. *Child Psychiatry and Human Development*, 22, 111-128.
- Chou, C., & Bentler, P. (1995). Estimates and tests in structural equation modeling. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and practice* (pp. 37-55). Thousand Oaks, CA: Sage.
- deHaas, P. A. (1986). Attentional styles and peer relationships of hyperactive and normal boys and girls. *Journal of Abnormal Child Psychology*, 14, 457-467.
- DuPaul, G. J., Power, T. J., Anastopoulos, A., Reid, R. (1998). *ADHD Rating Scale IV Manual*. New York: Guilford.
- DuPaul, G. J., Power, T. J., Anastopoulos, A., Reid, R., McGoey, K. E., & Ikeda, M. J. (1997). Teacher ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure, normative data, and psychometric properties. *Psychological Assessment*, 9, 436-444.
- DuPaul, G. J., Power, T. J., McGoey, K., Ikeda, M., & 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.
- Eme, R. F. (1992). Selective female affliction in the developmental disorders of childhood: A literature review. *Journal of Clinical Child Psychology*, 21, 354-364.
- Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology*, 26, 109-118.
- Epstein, M. A., Shaywitz, S. E., Shaywitz, B. A., & Woolston, J. L. (1991). The boundaries of attention deficit disorder. *Journal of Learning Disabilities*, 24, 78-86.
- Ernst, M., Liebenauer, L.L., King, A.C., Fitzgerald, G. A., Cohen, R.M., & Zametkin, A. J. (1994). Reduced brain metabolism in hyperactive girls. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 858-868.
- Faraone, S. V., Beiderman, J., Keenan, K., & Tsuang, M. T. (1991). A family-genetic study of girls with DSM-III attention deficit disorder. *American Journal of Psychiatry*, 148, 112-117.

- Frick, P. J., Lahey, B. B., Applegate, B., Kerdyck, L., Ollendick, T., Hynd, G. W., Garfinkel, B., Greenhill, L., Biederman, J., Barkley, R. A., McBurnett, K., Newcorn, J., & Waldman, I. (1994). DSM-IV field trials for the disruptive behavior disorders: Symptom utility estimates. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 529-539.
- Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1036-1045.
- Gordon, M. (1996). Must ADHD be an equal opportunity disorder? *ADHD Reports*, 4(2), 1-3.
- Gordon, M., & Mettelman, B. B. (1994). Gender differences in ADHD referrals: IQ, laboratory measures, and behavior ratings. *ADHD/Hyperactivity Newsletter*, 20, 7-9.
- Gualtieri, T., & Hicks, R. E. (1985). An immunoreactive theory of selective male affliction. *Behavioral and Brain Sciences*, 8, 427-441.
- Horn, W. F., Wagner, A. E., & Ialongo, N. (1989). Sex differences in school-aged children with pervasive attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 17, 109-125.
- Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin*, 95, 156-171.
- James, A., & Taylor, E. (1990). Sex differences in the hyperkinetic syndrome of childhood. *Journal of Child Psychology and Psychiatry*, 31, 437-446.
- Jarvinen, D. W., & Sprague, R. L. (1995). Using ACTeRS to screen minority children for ADHD: An examination of item bias. *Journal of Psychoeducational Assessment*, Special Issue on ADHD, 172-184.
- Joreskog, K., & Sorbom, D. (1993). *Manual for the LISREL 8*. Hillsdale, NJ: Erlbaum.
- Kashani, J., Chapel, J. L., Ellis, J., & Shekim, W. O. (1979). Hyperactive girls. *Journal of Operational Psychiatry*, 10, 145-148.
- 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., Kerdyck, L., Frick, P. J., Ollendick, T., Perez, D., Hart, E. L., Waldman, I., & Shaffer, D. (1994). DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*, 151, 1673-1685.
- Lambert, N.M., Sandoval, J., & Sassone, D. (1978). Prevalence of hyperactivity in elementary school children as a function of social system definers. *American Journal of Orthopsychiatry*, 48, 446-463.
- Loehlin, J. C. (1992). *Latent variable models*. Hillsdale, NJ: Erlbaum.
- McGee, R., Williams, S., & Silva, P. A. (1987). A comparison of girls and boys with teacher-identified problems of attention. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 711-717.
- Reid, R. (1995). Assessment of ADHD with culturally different groups: The use of behavior rating scales. *School Psychology Review*, 24, 537-560.
- Reid, R., DuPaul, G. J., Power, T. J., Anastopoulos, A.D., Rogers-Adkinson, D., Noll, M. B., & Riccio, C. A. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales. *Journal of Abnormal Child Psychology*, 26, 187-198.
- 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.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior assessment system for children*. Circle Pines, MN: American Guidance Service.

- Schachar, R., Sandberg, S., & Rutter, M. (1986). Agreement between teachers' ratings and observations of hyperactivity, inattention, and defiance. *Journal of Abnormal Child Psychology*, 14, 331-345.
- Seidman, L. J., Biederman, J., Faraone, S. V., Weber, W., Mennin, D., & Jones, J. (1997). A pilot study of neuropsychological function in girls with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 366-373.
- Silverthorn, P., Frick, P. J., Kuper, K., & Ott, J. (1996). Attention deficit hyperactivity disorder and sex: A test of two etiological models to explain the male predominance. *Journal of Clinical Child Psychology*, 25, 52-59.
- Sonuga-Barke, E., Minocha, K., Taylor, E., & Sandberg, S. (1993). Inter-ethnic bias in teachers' ratings of childhood hyperactivity. *British Journal of Developmental Psychology*, 11, 187-200.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Taylor, E. (1986). *The overactive child*. Philadelphia: Lippincott.
- 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.
- 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.
- Wechsler, D. (1974). *Wechsler intelligence scale for children-Revised*. New York: Psychological Corp.