Concordance of Self- and Informant Ratings of Adults’ Current and Childhood Attention-Deficit/Hyperactivity Disorder Symptoms

Marla Zucker, Mary K. Morris, Susan M. Ingram, Robin D. Morris, and Roger Bakeman
Georgia State University

Self-informant rating concordance for attention-deficit/hyperactivity disorder (ADHD) symptoms was assessed in 281 adults at the subscale (Inattention, Hyperactivity–Impulsivity) and individual symptom levels. Potential demographic, diagnostic, and informant identity moderators were also investigated. Concordance levels were similar for current and childhood symptoms. Although moderate positive correlations were found between self- and informant ratings on both subscales, informants endorsed more significant inattentive symptom severity. Kappa coefficients were variable, suggesting low concordance for certain symptoms. Sex and ADHD diagnosis moderated concordance, although effect sizes were small. These results have implications for the use of behavior rating scales in diagnosing ADHD, raise questions about the validity of self- and informant ratings, and support the need to investigate individual-differences variables that may impact concordance.

Childhood attention-deficit/hyperactivity disorder (ADHD) symptoms continue into adulthood for more than 50% of children with ADHD (Barkley, 1990). Follow-up studies of adolescents and adults who were diagnosed with ADHD as children, and of adults who present in clinical settings with significant inattention or hyperactivity–impulsivity (or both) with reported onset in early childhood, indicate that ADHD does not always end at adolescence (Biederman et al., 1994). Family genetic studies and a growing body of evidence for abnormalities in brain structure and function in both children and adults with ADHD further support the persistence of this disorder into adulthood (Biederman et al., 1994). However, there is no litmus test that can reliably determine the presence or absence of this disorder (Barkley, 2000, p. 2) in children or adults. Without such a test, diagnosis relies heavily on the use of behavior rating scales that permit comparison between the level of symptoms demonstrated by the individual being assessed and a normative sample (Mandal, Olmi, & Wilczynski, 1999; Reid & Maag, 1994).

Because of ADHD’s historical status as a disorder occurring in childhood, behavior rating scales for ADHD have been designed primarily for children and therefore emphasize parent and teacher report as the primary source of information, although some self-report measures for older children and adolescents are also available (e.g., Youth Self-Report; Achenbach, 1991b). Concordance between child-reported and adult-reported ADHD symptoms has been investigated for many different measures, although the Child Behavior Checklist (Achenbach, 1991a) and the Attention Deficit Disorder (ADD) module of the Diagnostic Interview for Children and Adolescents (Herjanic & Reich, 1982) have received the most attention. Although more substantive agreement between adult raters (i.e., parents and teachers) has been documented, concordance between child-reported and adult-reported symptoms is typically quite poor (e.g., Jensen et al., 1999; Loeber, Green, Lahey, & Soutphamer-Loeber, 1991). For example, Jensen et al. (1999) reported an average kappa value for parent–child ADHD diagnostic agreement of .19; Landis and Koch (1977) classify kappas of this magnitude as “slight.” Loeber et al. (1991) reported child-parent comparison rhos of .33 and .39; Cohen and Cohen (1983) regard correlation coefficients in the range of .30 to .50 as moderate. These findings of poor concordance between child and adult informants are consistent with similar studies of other childhood psychiatric disorders, particularly other externalizing disorders (e.g., Achenbach, McConaughy, & Howell, 1987; Stanger & Lewis, 1993). Reports by adult informants are also likely to carry more weight in diagnostic decision making for children (Jensen et al., 1999).

Unlike diagnosis of childhood ADHD, diagnosis of ADHD in adults relies heavily on information provided by the individual being assessed, either in a clinical interview or in responses to behavior rating scales. Reliance on adults’ self-report is further complicated by the diagnostic requirement that adults not only meet current criteria for the disorder, but also present with a history of clinically significant ADHD symptoms in childhood. Thus, diagnosing ADHD in adults presents a unique set of diagnostic challenges that has not been adequately addressed in the literature.

First, for adults who have not been diagnosed previously, establishing a retrospective diagnosis often requires the individual to recall and report his or her own childhood symptoms. The accuracy of retrospective self-reports of ADHD symptoms has been
questioned, as such self-reports are subject to inaccurate, incomplete, and distorted recall (Barkley, 2000; Shaffer, 1994). Second, it has been suggested that individuals with ADHD may have difficulty not only with accurately recalling their childhood symptoms, but also with reporting their current ADHD symptoms because of limited self-awareness of behaviors that have been present since early childhood (Wender, 1995). This difficulty with self-reflection and self-evaluation has been reported in anecdotal, clinical, and empirical literature on ADHD (e.g., Denckla, 1991; Hallowell & Ratey, 1994; Kugler & Kugler, 1997) and suggests that individuals with ADHD may not serve as the most valid informants of their own behavior.

To address the limitations of self-report, many researchers and clinicians enlist the help of a significant other, or informant, to corroborate and supplement the report of the patient before making a diagnosis of ADHD (Wender, 1995). However, this introduces another potential problem for diagnostic decision making: interobserver agreement. In previous studies, interobserver agreement on rating scales for ADHD symptoms has been measured in four different ways. First, correlations between observers’ overall ratings on symptom domain subscales (e.g., Inattention and Hyperactivity–Impulsivity) have been computed. Second, analyses have been conducted at the item level to assess observers’ agreement on individual ADHD symptoms. Third, interobserver level differences in subscale scores have been measured through t tests. Finally, studies have examined variations in diagnostic rates that would result from reliance on either self-report or informant data.

The level of agreement reported between self- and informants’ ratings of ADHD symptoms, both retrospectively and currently, has varied based on the method of analysis used. Significant correlations between adults’ and informants’ retrospective ratings of childhood ADHD symptoms have been reported at both global and symptom domain levels. Ward, Wender, and Reimherr (1993) assessed concordance between adults’ and parents’ retrospective ratings of ADHD symptoms in childhood; adults completed the Wender Utah Rating Scale, and their parents completed the Parents’ Rating Scale. Interrater correlations for the overall summary scores generated by these two measures were moderate but statistically significant for both a normative sample (r = .49) and for adults with ADHD (r = .41), even though these two measures differ in both number of items and item content. In a study with only a nonclinical sample (P. Murphy & Schachar, 2000), adults and informants both completed a rating scale of childhood ADHD symptoms based on the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM–IV; American Psychiatric Association, 1994), the ADHD Behavior Checklist for Adults (Barkley, 1995a). Interrater correlations were significant for both symptom domains represented on this measure (r = .76 for Hyperactive–Impulsive; r = .69 for Inattention) and were somewhat larger than those reported by Ward et al. (1993). However, significant differences in symptom severity were also noted, with adults rating themselves as having had more ADHD symptoms as children than did their parents. Overall, the results of these studies indicate that adults may be able to provide some retrospective data regarding their childhood ADHD symptoms that concur with data provided by their parents, although there is some suggestion that, at least in a nonclinical sample, they may overestimate the severity of their childhood symptoms.

The results of studies investigating self–informant concordance in ratings of adolescents’ and adults’ current ADHD symptoms are mixed, although concordance rates are somewhat higher than those reported between younger children and their parents. Two follow-up studies assessing interrater agreement between adolescents and young adults diagnosed with ADHD in childhood and their parents suggest that informants report a more significant level of symptoms. Parents were more likely to report inattentiveness, hyperactivity, and impulsivity in their adolescents than were the adolescents themselves in an interview tailored to the DSM–III (American Psychiatric Association, 1980) criteria for attention deficit disorder with hyperactivity (ADD–H; Gittelman & Mannuzza, 1985). In addition, 21 of the adolescents in this study met DSM–III diagnostic criteria for full or partial ADD–H on the basis of parent report, but not self-report, of symptoms. In a similar manner, in a young adult follow-up study, 28% of the participants rated themselves as 1.5 standard deviations above the mean on a DSM–III–R-based (American Psychiatric Association, 1987) ADHD rating scale, whereas 58% met this criteria based on parent ratings on a parallel checklist (Barkley, 1997). In a nonclinical sample, adult and informant ratings of current symptoms on the ADHD Behavior Checklist were significantly correlated (Hyperactive–Impulsive subscale r = .59; Inattention subscale r = .70) but did not differ in level of severity, in contrast to findings for retrospective reports in this sample (P. Murphy & Schachar, 2000).

Only one previous study has assessed self–informant concordance for ratings of individual symptoms in adults with ADHD (Downey, Stelson, Pomerleau, & Giordani, 1997). When assessed as percentage agreement regarding the presence or absence of each symptom, item-level concordance between adult and informant ratings for current DSM–III–R ADHD symptoms on the Symptom Checklist for ADHD Adults was high (range = 57.8% to 89.1%). However, a more stringent measure of interobserver agreement that takes into account chance agreement suggested only poor to moderate interobserver agreement for the presence or absence of individual ADHD symptoms (kappa range = −.08 to .50).

The problem of discordant self- and informant ratings of childhood and current ADHD symptoms poses a difficult challenge to diagnosing ADHD in adults. Discrepancies between patients and their informants can lead to a diagnostic dilemma because there is no gold standard to which to compare their ratings, and investigators have reached different conclusions regarding the validity of self-reported ADHD symptoms in both adolescents and adults (Barkley, 1997; Biederman et al., 1993; Mannuzza, Klein, Bessler, Malloy, & LaPadula, 1993). A review of previous studies suggests that the time frame of rating (current vs. retrospective), the level of analysis (item vs. subscale), and the sample (ADHD vs. nonclinical) may all influence interrater reliability.

The present study investigated the degree of agreement between adults’ and informants’ ratings of both childhood (retrospective) and current ADHD symptoms on the ADHD Behavior Checklist for Adults (Barkley, 1995a) in a large sample of college students who presented to an on-campus clinic with academic difficulties and who subsequently received a comprehensive evaluation. Self- and significant other ratings on this checklist were collected as part of this clinical assessment. On the basis of the results of this evaluation, a subset of this sample was diagnosed with ADHD, consistent with DSM–IV criteria. This study expands previous research through its larger sample size and, thus, greater power to
detect differences; assessment of both childhood and current symptoms; assessment of concordance using several statistical methodologies at both global (subscale) and item levels; and investigation of demographic (age and sex), diagnostic (ADHD and learning disability), and informant variables as potential moderators of concordance.

Method

Participants

Participants were selected from a database containing the results of psychoeducational evaluations of college students who presented with academic difficulties to an on-campus assessment clinic between 1993 and 2001 (n = 387). This clinic receives referrals from disability service providers based at 15 colleges and universities in the state university system, including both 2-year and 4-year institutions. Evaluations were conducted to determine the presence of learning disabilities, ADHD, and psychiatric disorders and to recommend appropriate educational accommodations. The evaluations included the following: (a) self-report questionnaire and computerized measures of past and current social–emotional functioning (Symptom Checklist–Revised [SCL–90–R], Derogatis, 1992; Structured Clinical Interview for DSM–IV Screen, First, Gibbon, Williams, & Spitzer, 1996) to screen for a broad range of psychiatric symptoms and to identify possible DSM–IV Axis I diagnoses; (b) self-rating of both childhood and current symptoms of ADHD on the ADHD Behavior Checklist for Adults; (c) a semi-structured clinical interview in which detailed information regarding developmental, educational, medical, social, and vocational history was obtained and responses to the self-report measures described above were reviewed and clarified to establish the presence or absence of possible DSM–IV Axis I diagnoses; (d) standardized intellectual assessment; (e) academic achievement testing; (f) a neuropsychological assessment battery assessing a broad range of cognitive domains, including auditory–phonological awareness, language, visual–perceptual–motor abilities, attention, memory, executive functions, and fine motor skills; and (g) behavioral observations gathered throughout the interview and standardized testing procedures. Informant report versions of the ADHD Behavior Checklist for Adults for both childhood and current symptoms of ADHD were also collected as part of the standard evaluation.

Results of each individualized assessment were reviewed and integrated by an assessment team, led by a licensed clinical psychologist. Both self- and informant ratings on the ADHD Behavior Checklist for Adults were available to the team and informed the diagnostic decision-making process; however, these data were integrated with all other data generated in the evaluation and were not used in isolation to make the diagnosis of ADHD. The absence of clinically significant symptom ratings by either self or informant did not preclude a diagnosis of ADHD if, based on expert clinical judgment, the preponderance of other data supported both symptom presence and significant functional impairment. Participants were diagnosed with ADHD or other Axis I psychiatric disorders if they were judged to fully meet either DSM–III–R (1993–1994) or DSM–IV (1994–2001) criteria for these disorders. Thus, participants diagnosed with ADHD (a) exhibited six of nine possible symptoms of either inattention or hyperactivity–impulsivity or both; (b) evidenced onset of these symptoms in childhood; (c) experienced significant academic, vocational, or social impairment as a result of these symptoms that was present in more than one specific environmental situation; and (d) did not present with other psychiatric disorders (e.g., mood disorder, substance abuse) that might better account for the reported symptoms.

The assessment team diagnosed students with a learning disability in math, reading, or written expression if they met the following criteria: (a) at least average intellectual ability (either Verbal or Nonverbal IQ Score ≥ 90), (b) a significant discrepancy between intellectual and academic achievement, (c) documentation of cognitive processing deficit(s) linked to the academic deficit, and (d) academic and cognitive deficits are not better accounted for by social–emotional factors.

Participants were selected if both self- and informant ratings on checklists of current and childhood ADHD symptoms were available (n = 281). Individuals with one or more missing data points on any of the ADHD questionnaires (self or informant) were excluded (n = 106). The group excluded because of missing data was significantly older than the participant group, t(385) = −4.46, p < .01, but was comparable in sex, ethnicity, Full Scale IQ scores, and prevalence of diagnoses of learning disability, ADHD, and other psychiatric disorders.

Demographic characteristics of the sample are presented in Table 1. The sample was 53.7% male (n = 151) and 46.3% female (n = 130). The mean age of the sample was 23.59 years (SD = 6.60), and mean Full Scale IQ score was 95.42 (SD = 10.25). Racial–ethnic composition was 84.7% Caucasian, 7.8% African American, and 7.5% of other ethnicity. Of the sample, 29.2% were diagnosed with one or more Axis I psychiatric disorders (19.2% mood disorders, 10.7% anxiety disorders, and 4.6% other psychiatric disorders). Developmental learning disabilities were diagnosed in 65.5% of the sample, ADHD in 34.2%, neurological illness or injury in 7.1%, and other medical disorders in 2.5%. Twelve and a half percent of the sample received no diagnosis.

As shown in Table 2, the overwhelming majority of informants for the retrospective report of symptoms were parents (92.5%), and parents comprised 69.8% of the informants for the current report of symptoms. Informants for current report of symptoms were also friends or roommates (13.2%), spouses or partners (10.7%), and other relatives and relationships (5.7%).

Measure

Participants and informants completed two versions of the ADHD Behavior Checklist for Adults ( Barkley, 1995a). This checklist lists the specific inattentive and hyperactive–impulsive symptoms that compose the DSM–IV criteria for ADHD. The first version assessed childhood symptoms, and the second version assessed symptoms during the past 6 months.

Table 1

<table>
<thead>
<tr>
<th>Demographic Characteristics of the Sample (N = 281)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
</tr>
<tr>
<td>Caucasian</td>
</tr>
<tr>
<td>African American</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>Full Scale IQ score</strong></td>
</tr>
<tr>
<td><strong>Diagnoses</strong></td>
</tr>
<tr>
<td>Learning disability</td>
</tr>
<tr>
<td>ADHD</td>
</tr>
<tr>
<td>Any psychiatric disorder</td>
</tr>
<tr>
<td>Mood disorder</td>
</tr>
<tr>
<td>Anxiety disorder</td>
</tr>
<tr>
<td>Other psychiatric</td>
</tr>
<tr>
<td>Neurological disorder</td>
</tr>
<tr>
<td>Medical disorder</td>
</tr>
<tr>
<td>No diagnosis</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit/hyperactivity disorder.

*IQ was assessed by the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981; n = 109) or by the Wechsler Adult Intelligence Scale—III (Wechsler, 1997; n = 163). *a Multiple diagnoses account for percentages totaling more than 100.
checklists were completed at the same time, in reference to the same
represent independent data points, which is not surprising given that both
mants (.65 for Hyperactive–Impulsive subscales). Given the significant, but only
moderate, correlations between the Inattention and Hyperactive–Impulsive subscales (mean \( r = .61 \)), we conducted subsequent analyses separately for each scale. Because of the redundancy between summary scores and
symptom counts within each subscale (mean \( r = .97 \)), subsequent analyses
used only summary scores. Cronbach’s alpha was used to assess the internal reliability of the Inattention and Hyperactive–Impulsive subscales. We computed alphas separately for self- and informant ratings.

To assess the relation between self- and informant ratings, we used three methods. First, we performed Pearson product–moment correlations to assess the relation between self- and informant ratings at the subscale level. Second, we computed weighted kappa coefficients to assess acrossrater agreement at an item level. Kappa weights were assigned to reflect agreement or disagreement in evaluating whether symptoms were present at a clinically significant level (i.e., rating of “often” or “very often”). Any rating that crossed this boundary of clinical significance (i.e., ratings of 0 or 1 vs. ratings of 2 or 3) was considered a disagreement, and ratings that did not cross this boundary were counted as agreements. Therefore, discrepancies of one, two, or three points could be considered disagreements under this system. Finally, we used repeated measures analysis of variance (ANOVA) to assess the direction, significance, and effect size of mean differences between self- and informant ratings for subscale scores.

To investigate whether demographic characteristics (sex and age), par
ticipant diagnoses (ADHD and learning disability), or informant identity
moderated concordance between self- and informant report of ADHD
symptoms, we performed hierarchical multiple regressions. Participants’ self-ratings were first regressed on informants’ ratings and one of the demographic, diagnostic, or informant variables and then on a variable representing the interaction between the informant’s rating and the poten
tial moderator. To reduce possible problems with multicollinearity, interval-scaled variables were centered (i.e., means were subtracted from raw scores; Cronbach, 1987) before interaction terms were computed.

The percentage of participants whose symptom ratings exceeded a
normative cutoff criterion based on self-ratings, informant ratings, or both
ratings was compared using Pearson chi-square analyses. Cutoff criteria for both self- and informant ratings were based on the norms available for self-report of current and childhood symptoms, as no norms are available for informant report (K. Murphy & Barkley, 1995). Scores greater than 1.5 standard deviations above the mean based on age for current symptoms, and age and sex for childhood symptoms, were considered clinically significant. Finally, concordance between rating scale data and clinician
assigned diagnosis of ADHD, based on a large set of clinical data that
included these ratings, was also explored. Although multiple strategies
could have been used in this comparison, we chose to use the most
stringent approach, requiring that all four sets of ratings reflect a significant level of symptoms. Thus, in order to classify a participant in the ADHD
group based solely on checklist data, we required that symptom severity in either the Inattention or Hyperactive–Impulsive domain (or both) meet a
norms-based cutoff criterion (\( \geq 1.5 \ SD \)) for both childhood and current
symptoms and based on both self- and informant ratings.

Results

Reliability

The internal reliability coefficients (Cronbach’s alpha) for the
ADHD Behavior Checklist for Adults ranged from .89 to .91 for the
Inattention subscale and from .84 to .91 for the Hyperactive–
Impulsive subscale. These alphas suggested excellent internal reliabil
ity for both self- and informant ratings of childhood and current
symptoms.
Self–Informant Concordance

We computed correlation coefficients between self- and informant summary scores. As shown in Table 3, analyses revealed moderate but significant (p < .01) positive correlations between self- and informant ratings of current symptoms on both the Inattention and Hyperactive–Impulsive subscales. Similar correlations were evident for childhood symptoms.

Table 4 shows the weighted kappa coefficients yielded for each item on the ADHD Behavior Checklist. The kappas for individual items indicated “slight” to “moderate” interrater concordance according to the classification system of Landis and Koch (Landis & Koch, 1977). Mean kappas for both current symptoms and childhood symptoms were similar, indicating a “fair” level of agreement. Similar levels of item concordance were evident for both the Inattention and Hyperactive–Impulsive subscales. Given the method by which kappa weights were assigned, this finding suggests that concordance at the item level was fair, even when raters were indicating only whether or not ADHD symptoms were present at a clinically significant level.

Table 5 presents descriptive statistics and results of repeated measures ANOVAs for self- and informant ratings of current and childhood symptoms. Informants rated participants as showing significantly more current inattention than did participants themselves, whereas self- and informant ratings of current hyperactive–impulsive symptoms did not differ significantly. The same pattern of results was present for childhood symptoms and remained the same when analyses were repeated separately in the ADHD and non-ADHD subgroups.

Moderating Effects of Demographic, Diagnostic, and Informant Variables

Table 6 presents results of hierarchical regression analyses that tested whether the demographic, diagnostic, or informant variables moderated the association between self- and informant ratings. In each case, the informant rating accounted for the largest proportion of variance. Although no predictions were made regarding significant main effects of the potential moderating variables, several were found. ADHD and learning disability status alone accounted for significant variance in current self-ratings; for both symptom types, adults with ADHD reported more symptoms, whereas adults with learning disabilities reported fewer symptoms. In addition, a main effect for age was found for current self-ratings, with older adults reporting more inattention symptoms. Similar main effects of ADHD diagnosis and age were found for childhood symptoms. The ADHD group reported more symptoms of inattention and hyperactivity–impulsivity, and older adults reported more inattention symptoms in childhood.

As indicated in Table 6, some significant interaction effects were found, consistent with a moderating effect, but the effect sizes were quite small, accounting for 1% or less of the variance. For ratings of childhood behavior, both sex and ADHD diagnosis, but not age or learning disability, moderated the relation between self- and informant ratings of inattention and hyperactive–impulsive symptoms. To be specific, concordance between self- and informant childhood ratings was stronger for females than for males for both types of symptoms. The moderating effect of ADHD diagnosis was dependent on symptom domain; for ratings of childhood inattention, participants with ADHD were less concordant with their informants, whereas for childhood symptoms of hyperactivity–impulsivity, concordance was stronger for participants with ADHD. For ratings of current symptoms, possible moderating effects of ADHD diagnosis were found only for inattentive symptoms, similar to childhood ratings. Informant identity (parent vs. non-parent) did not moderate concordance for current report of symptoms. The small number of non-parent informants for childhood symptoms precluded analysis of informant identity as a moderator for report of childhood symptoms.

Normative Criteria for Symptom Severity: Comparisons Across Informants and With Clinician Diagnosis

Table 7 shows the percentage of participants who would have met norms-based criteria for a significant level of symptoms according to self-, informant, and both self- and informant report on the checklist. Pearson chi-square analyses revealed that informants’ ratings of current ADHD symptoms (Inattention, Hyperactive–Impulsive, and Total) were significantly more likely to meet cutoff criteria than were self-ratings. The same pattern of results was found for ratings of childhood symptoms.

Assignment to an ADHD group, based on applying a cutoff criterion for significance to self- and informant ratings, was also compared to clinician-assigned diagnoses of ADHD, based on the integrated results of a comprehensive evaluation. Because clinicians had access to the rating scale data in the diagnostic process, assignment to groups in these two conditions was not completely independent. As shown in Table 8, clinician diagnosis and checklist data were concordant for 71% of participants (n = 200). Among the 81 participants for whom checklist scores and clinician diagnosis were discordant, the majority (n = 75) received ADHD diagnoses but failed to meet the stringent cutoff criteria for the checklist data. Further exploration of these 75 participants revealed that approximately half of this group (47%) met cutoff criteria for both childhood and current symptoms according to one informant but not the other (24 based on informant report only; 11 based on self-report only). However, the remainder (53%) did not meet cutoff criteria for both childhood and current symptoms based on a single rater. Of these 40 adults who received clinician diagnoses of ADHD, the majority (85%) presented with a significant level of symptom severity on at least one rating scale (either self- or informant report for either childhood or current symptoms). Furthermore, all but three of these individuals had at least six current and at least six childhood inattention or hyperactive–impulsive symptoms endorsed at some level (1–3). Clinicians’ decisions to reject a diagnosis of ADHD when presented with significant ratings from both the participant and the informant were an infrequent event (n = 6); the majority of these participants were
assigned another DSM–IV diagnosis by the clinical team (e.g., mood or anxiety disorder or substance abuse) on the basis of a judgment that the reported symptoms were better accounted for by another disorder.

Discussion

The present study assessed concordance of self- and informant ratings of childhood and current ADHD symptoms in a sample of college students referred for clinical evaluation because of academic difficulty. On the basis of the results of this evaluation, a subset of this sample received a clinical diagnosis of ADHD.

Self-informant concordance on the ADHD Behavior Checklist for Adults was assessed at the subscale and individual symptom levels, and demographic, diagnostic, and informant variables were explored as potential moderators of concordance. Finally, the proportion of individuals who met a norms-based cutoff criterion for significant symptom severity was compared across informants and with clinician diagnoses.

For current symptoms, correlations between self- and informant ratings for both Inattention and Hyperactive–Impulsive subscales were significant but moderate, consistent with previous findings based on this checklist in a nonclinical sample (P. Murphy & Schachar, 2000). In a similar manner, the mean item-level concor-
dance for current symptoms was also moderate and appeared to be similar for both Inattention and Hyperactive–Impulsive items. However, kappa values ranged from slight to moderate, indicating a wide range of agreement on individual symptoms, even when evaluating only whether ADHD symptoms were present at a clinically significant level. The mean kappa level for DSM–IV ADHD symptoms found in this study was somewhat higher than the concordance reported by Downey et al. (1997) for DSM–III-R symptoms on the Symptom Checklist for ADHD Adults. However, those authors noted that their finding of poor agreement may have been due to the skewed distribution of symptoms in their clinical sample of adults with ADHD. The increased item-level concordance in our sample may be due in part to the broader distribution of symptom ratings in our more diverse clinical sample.

Table 6
Hierarchical Regression Model Assessing Demographic, Diagnostic, and Informant Identity Moderators of Self–Informant Concordance

<table>
<thead>
<tr>
<th>Potential moderator</th>
<th>Step 1a</th>
<th></th>
<th></th>
<th>Step 2a</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta R^2$</td>
<td>$\beta 1^b$</td>
<td>$\beta 2^b$</td>
<td>$\Delta R^2$</td>
<td>$\beta 3^b$</td>
</tr>
<tr>
<td><strong>Current symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.31**</td>
<td>.10*</td>
<td>.55**</td>
<td>.00</td>
<td>−.04</td>
</tr>
<tr>
<td>Sex</td>
<td>.30**</td>
<td>.05</td>
<td>.54**</td>
<td>.00</td>
<td>−.08</td>
</tr>
<tr>
<td>ADHD</td>
<td>.33**</td>
<td>.21**</td>
<td>.55**</td>
<td>.01*</td>
<td>−.18*</td>
</tr>
<tr>
<td>Learning disability</td>
<td>.31**</td>
<td>−.13*</td>
<td>.52**</td>
<td>.00</td>
<td>.03</td>
</tr>
<tr>
<td>Informant identity</td>
<td>.30**</td>
<td>−.06</td>
<td>.54**</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Hyperactive–Impulsive subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.33**</td>
<td>.06</td>
<td>.57**</td>
<td>.00</td>
<td>.06</td>
</tr>
<tr>
<td>Sex</td>
<td>.33**</td>
<td>−.02</td>
<td>.57**</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>ADHD</td>
<td>.34**</td>
<td>.12*</td>
<td>.53**</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Learning disability</td>
<td>.34**</td>
<td>−.11*</td>
<td>.56**</td>
<td>.00</td>
<td>−.09</td>
</tr>
<tr>
<td>Informant identity</td>
<td>.33**</td>
<td>.01</td>
<td>.57**</td>
<td>.00</td>
<td>−.04</td>
</tr>
<tr>
<td><strong>Childhood symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.33**</td>
<td>.11*</td>
<td>.57**</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Sex</td>
<td>.32**</td>
<td>.05</td>
<td>.55**</td>
<td>.01*</td>
<td>−.16*</td>
</tr>
<tr>
<td>ADHD</td>
<td>.37**</td>
<td>.28**</td>
<td>.41</td>
<td>.01*</td>
<td>−.15*</td>
</tr>
<tr>
<td>Learning disability</td>
<td>.32**</td>
<td>−.06</td>
<td>.56**</td>
<td>.00</td>
<td>−.01</td>
</tr>
<tr>
<td><strong>Hyperactive–Impulsive subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.42**</td>
<td>.05</td>
<td>.65**</td>
<td>.01</td>
<td>.08</td>
</tr>
<tr>
<td>Sex</td>
<td>.42**</td>
<td>−.02</td>
<td>.65**</td>
<td>.01*</td>
<td>−.14*</td>
</tr>
<tr>
<td>ADHD</td>
<td>.45**</td>
<td>.19**</td>
<td>.57**</td>
<td>.01*</td>
<td>.14*</td>
</tr>
<tr>
<td>Learning disability</td>
<td>.42**</td>
<td>.02</td>
<td>.65**</td>
<td>.00</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit/hyperactivity disorder.

*a* Step 1: informant rating and potential moderator; *Step 2*: Informant Rating × Potential Moderator.  
$^b$ $\beta 1$: main effect of the variable of interest; $\beta 2$: main effect of informant rating, which was entered in every model; $\beta 3$: beta for the interaction.  
$^c$ Parent vs. nonparent.

$p < .05. \ \ \ \ \ \ \ ** p < .01.$

Table 7
Percentage of Participants Meeting Norms-Based Criteria for Significant Symptoms by Rating Source (N = 281)

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Source</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self</td>
<td>Informant</td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td><strong>Current symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention</td>
<td>28.1</td>
<td>38.4</td>
<td>18.1</td>
<td></td>
</tr>
<tr>
<td>Hyperactive–Impulsive</td>
<td>11.0</td>
<td>15.3</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>18.1</td>
<td>24.9</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td><strong>Childhood symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattention</td>
<td>17.4</td>
<td>27.4</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>Hyperactive–Impulsive</td>
<td>11.4</td>
<td>13.2</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>15.7</td>
<td>17.4</td>
<td>8.5</td>
<td></td>
</tr>
</tbody>
</table>

Note. Chi-squares of self–informant comparisons for all scales were significant ($p < .01$).
Although self- and informant ratings of current symptoms were positively correlated, the level of symptom severity reported differed between participants and informants. Informant ratings of current inattention were significantly higher, whereas ratings of hyperactivity–impulsivity were similar. This pattern was found for both the ADHD and non-ADHD participants, suggesting that it does not reflect a unique feature of adults with ADHD. Gittelman and Mannuzza (1985) found a similar pattern in a sample of adolescents with ADHD, although parents in this study also rated their adolescents higher in hyperactivity and impulsivity. In contrast, when the ADHD Behavior Checklist was administered to a nonclinical sample, P. Murphy and Schachar (2000) found no mean differences between self- and informant ratings of current symptoms for either domain.

As would be expected from these mean differences, more adults met normative criteria for significant levels of current ADHD symptoms—particularly within the inattention domain—based on informant report than on self-report. However, it must be noted that cutoff scores for informants were determined using published norms for self-report, as no normative scores are available for informant report (K. Murphy & Barkley, 1995). These findings support the need for separate normative data for self- and informant ratings for adults. It remains unclear whether young adults having academic difficulty experience a diminished awareness or reluctance to acknowledge problems with inattention or whether their significant others may be biased toward increased symptom reporting, perhaps reflecting concern about the difficulties that led to the participant’s referral for evaluation.

Self–informant correlations at the subscale level for ratings of childhood symptoms were also moderate and consistent with those reported for this measure in a nonclinical sample of adults (P. Murphy & Schachar, 2000) and with similar measures for both a nonclinical sample and for adults with ADHD (Ward et al., 1993). Interobserver concordances for childhood symptoms at the item level, which had not been previously evaluated, were consistent with those for current symptoms in this sample. Kappa values ranged from fair to moderate and were similar for symptoms in both the inattention and hyperactive–impulsive domains. Therefore, despite the specific concern raised in the literature regarding the potential inaccuracy of retrospective self-reports of ADHD symptoms (Barkley, 2000; Shaffer, 1994), rater concordance for retrospective report does not appear to differ from that of current report.

Just as with current symptoms, informants rated participants higher in childhood inattention but not hyperactivity–impulsivity. Mean level differences had not been previously explored in a clinical sample; however, within a nonclinical sample, adults retrospectively rated themselves as more inattentive and hyperactive–impulsive in childhood than did their informants on this checklist (P. Murphy & Schachar, 2000). This disparity in results between clinical and nonclinical samples does not appear to be related to specific characteristics of ADHD because for both childhood and current symptoms, this same pattern of findings was present in both those participants diagnosed with ADHD and those who were not. Consistent with findings for current symptoms, more adults in this study met normative criteria for significant childhood symptoms of ADHD by their informants’ report than by their own report. Again, this discrepancy was driven primarily by ratings on the Inattention scale.

Exploration of potential moderators of the level of rater concordance resulted in some statistically significant findings; however, the small effect sizes for these moderators necessitate cautious interpretation. Neither age nor learning disability status were found to be significant moderators of concordance. The former finding is not unexpected given the limited age range of our sample and is consistent with results reported by P. Murphy and Schachar (2000), who found similar correlations between self- and informant ratings across age groups. Informant identity also did not serve as a moderator of concordance, although this could only be assessed for ratings of current symptoms because the overwhelming majority of informants for childhood symptoms were parents.

There was some suggestion that female participants were more concordant with their informants when rating their childhood symptoms of both inattention and hyperactivity–impulsivity. A difference in the socialization process for boys and girls may provide one possible explanation for this finding. Previous studies suggest that parents are more elaborative, use more emotion words, and provide more detailed evaluations and recollections when talking with daughters about past experiences and that girls tend to recall more detailed and coherent information about the past, particularly the emotional aspects of past events (e.g., Fivush, 1998). These childhood socialization experiences and capabilities may enhance adult females’ retrospective recall of their childhood behavior.

ADHD diagnostic status also appeared to exert some limited impact on concordance. For participants with ADHD, concordance with informants was lower for inattentive symptoms, for both childhood and current ratings. This finding is consistent with clinical descriptions of difficulty with accurate self-evaluation in individuals with ADHD (e.g., Denckla, 1991; Hallowell & Ratey, 1994; Kugler & Kugler, 1997), but the accuracy of informant ratings can also be questioned. In contrast, for hyperactive–impulsive symptoms, concordance was higher in the ADHD group, although this was only evident for ratings of childhood symptoms. This finding may be related to the problematic and disruptive nature of children’s hyperactivity and impulsivity and the attention that parents and teachers may pay to such behaviors as a result. That is, there may have been high levels of communication about children’s hyperactive–impulsive behaviors in families of children with ADHD, which could lead to greater concordance in retrospective ratings of these childhood behaviors.

These results highlight the importance of the level of analysis used to assess interobserver agreement on the ADHD Behavior Checklist for Adults and on other behavior rating scales for ADHD with a similar format. Although concordance is moderate when assessed at a subscale level, assessment of agreement on individual items suggests considerable variability, with fairly low concordance for certain symptoms. Both subscale- and item-level approaches to interpreting checklist data are used in clinical practice. A more global approach to checklist data sums individual ratings across a set of symptoms to create a continuous variable, which can then be compared to a cutoff criterion for significant deviation from normative data (e.g., K. Murphy & Barkley, 1995). An alternate strategy focuses on establishing the presence or absence of six or more inattentive or hyperactive–impulsive symptoms at a designated level of symptom severity, representing an item-level or so-called symptom count method of using checklist data. Valid and clinically useful estimates of interrater reliability must be
RATING CONCORDANCE FOR ADHD SYMPTOMS

387

based on the way in which checklist data will actually be used in the diagnostic process.

Correlational analyses have been used in several previous studies to estimate interrater reliability on behavioral checklists of ADHD symptoms. However, this type of analysis is not sensitive to systematic differences in the level of symptom severity reported across raters. The fact that informants endorsed higher levels of inattentive symptoms than participants, for both current and childhood ratings, raises the possibility of a systematic rater bias that negatively impacts interrater reliability on this measure. It is not clear why this finding did not extend to ratings of hyperactivity—impulsivity. The inattentive and hyperactive behaviors included on this checklist may differ in how readily they can be observed and judged by others, as has been suggested for undercontrolled versus overcontrolled problems (Achenbach et al., 1987). For example, behaviors such as “driven by a motor” or “talking excessively” might be more easily identified by informants than behaviors such as “difficulty sustaining attention in tasks” or “easily distracted.” However, this hypothesis is not supported by an examination of kappa values for individual symptoms of inattention and hyperactivity—impulsivity. Further investigation of the properties of specific items that might contribute to lower levels of Interrater reliability is needed. For example, it has been suggested that the DSM–IV thresholds for diagnosing ADHD in adults, particularly for inattentive symptoms, may need to be lowered to reflect developmental changes in ADHD symptomatology (DuPaul et al., 2001; Faraone, Biederman, Feighner, & Monuteaux, 2000; Heliggenstein, Conyers, Berns, & Smith, 1998). If the DSM–IV Inattention criteria are developmentally insensitive, this could negatively impact interrater reliability for this symptom domain.

This systematic difference between raters also raises critical questions about the relative validity of self- and informant ratings of ADHD symptoms. Without a clear criterion measure or gold standard to which to compare these ratings, the validity of self- and informant ratings on checklists of ADHD symptoms cannot be definitively established. Because of the subjective nature of behavior ratings, two raters could observe the same behavior yet rate its intensity, duration, and frequency differently (Mandal et al., 1999; Reid & Maag, 1994). Discrepancies in ratings could reflect raters’ differing attitudes and judgments, interpretations of items, scales and behavior, biases toward the individual being rated, and levels of tolerance for various behaviors (Barkley, 1995b; Mandal et al., 1999; Reid & Maag, 1994). Second, different settings place different demands on a person’s behavior, and thus discrepant ratings can also reflect differences in the person’s behavior across settings (Barkley, 1995b; Mandal et al., 1999). Although adults have a window on their own behavior at all times and across all contexts, their informants observe their behavior only in certain contexts. Therefore, cross-situational and cross-setting variability can be another confounding factor for interobserver concordance. Third, retrospective reports present an additional complication. The task of responding to a questionnaire about ADHD symptoms experienced in childhood does not draw on one’s memory of a single episode or salient life event. Memory for such experiences is classified as episodic (Kessler, Mroczek, & Belli, 1999), and adults’ episodic memory for the main features of a salient childhood experience has been shown to be generally accurate and stable over time (Brewin, Andrews, & Gotlib, 1993). Instead, memory for ADHD symptoms experienced in childhood is organized more as semantic memory, a general memory for a prototypical experience of a repeated and frequent symptom (Kessler et al., 1999). Although there is some evidence that questionnaire-based reports that draw on semantic memories can provide accurate information about the salient and constant aspects of repeated experiences and behaviors (for a review, see Brewin et al., 1993; Kessler et al., 1999), the more constructed nature of such memories may also impact their validity.

In the absence of strong support for the interrater reliability and criterion validity of behavioral rating scales designed to assess ADHD in adults, clinicians have little empirical evidence to guide them in using these instruments as part of the diagnostic process. Although there is apparently widespread agreement that these types of scales cannot be used in isolation to establish a diagnosis of ADHD, it remains unclear how strongly rating scale data should be weighted in the decision-making process and how it can best be integrated with other types of data to enhance diagnostic accuracy. In particular, the impact of discordant ratings provided by the client and by significant others on clinician judgments is not known, although such disagreements are a frequent experience in the evaluation process.

Although clinician-generated diagnoses of ADHD were available for this sample, they could not be used as a criterion measure of validity because they were not completely independent of the rating scale data. Nonetheless, a comparison of concordance between a significant level of symptom endorsement on the checklist and a clinical diagnosis may be valuable as a source of hypotheses about the ways in which clinicians use this type of data. It is not surprising that this comparison suggests that self- and informant concordance in rating both current and childhood symptoms as clinically significant on this particular checklist was associated with a high probability of a clinician diagnosis of ADHD. In this sample, it is also clear that clinicians frequently assigned a diagnosis of ADHD when stringent criteria of endorsement of both childhood and current symptoms by two raters were not met. Further exploration of these individual cases did not suggest a systematic valuation of one rating source by clinicians. Rather, in making diagnostic decisions, clinicians most likely weigh data from different sources on the basis of information specific to a given client. That is, clinicians are likely to make judgments about both clients’ and informants’ capability and probability of providing accurate information—based on a variety of individual and contextual factors—and assign weights to their ratings accordingly. In addition, given the questionable reliability of specific ratings on a Likert scale, clinicians may also use their clinical judgment in diagnosing ADHD when six or more symptoms of inattention or hyperactivity—impulsivity are endorsed, but some are not at a clinically significant level of severity.

Possible limitations of the present study deserve comment. The degree to which these results can be extended to the general adult population remains unknown given that the current sample was composed of clinic-referred college students. In a similar manner, this study focused on one particular DSM–IV-based checklist of ADHD symptoms; results may not be generalizable to all checklists. In addition, although this study addressed findings for childhood and current symptoms separately, measures of childhood and current symptoms cannot be considered independent observations, which limits the conclusions that might be drawn from their direct comparison. Finally, because the clinician diagnoses in this study
were not based on structured interviews, their reliability cannot be evaluated. This lack of reliability data, as well as the nonindependence of diagnoses from rating scale data, prevented us from using clinician diagnoses of ADHD to address questions of criterion validity in this study.

In summary, these results provide further data regarding the reliability of a frequently used behavior rating scale for ADHD symptoms in a diverse clinical sample of adults with academic difficulties. Concordance of self- and informant ratings was moderate at best but similar for both retrospective and current ratings and for both inattentive and hyperactive–impulsive symptoms. Reliability estimates varied depending on the level of analysis (i.e., subscale vs. individual items), and a systematic difference between raters was evident for both childhood and current ratings of inattention, with informants endorsing more significant symptom severity. Both sex and the presence of an ADHD diagnosis contributed a small but statistically significant moderating effect on the level of concordance between self- and informant ratings. Empirical support for the validity of self- and informant reports of ADHD symptoms on behavioral rating scales remains to be established and is impeded by the absence of a clear criterion measure. Further exploration of the contribution and potential incremental validity of these types of rating scales in the assignment of clinical diagnoses of ADHD in adults is needed.

References

New Editors Appointed, 2004–2009

The Publications and Communications Board of the American Psychological Association announces the appointment of five new editors for 6-year terms beginning in 2004.

Except where noted, as of January 1, 2003, manuscripts should be directed to the following individuals:

- For Psychology and Aging (http://www.apa.org/journals/pag.html), submit manuscripts to Rose T. Zacks, PhD, Department of Psychology, Michigan State University, East Lansing, MI 48824-1117.
- For Psychological Assessment (http://www.apa.org/journals/pas.html), submit manuscripts to Milton E. Strauss, PhD, Department of Psychology, Case Western Reserve University, Cleveland, OH 44106-7123.
- For Journal of Family Psychology (http://www.apa.org/journals/fam.html), submit manuscripts to Anne Kazak, PhD, ABPP, Oncology Psychosocial Research, The Children’s Hospital of Philadelphia, Room 1486 (Market Street), 34th and Civic Center Boulevard, Philadelphia, PA 19104. For overnight couriers: Room 1486, 3555 Market Street, Philadelphia, PA 19104.
- For Journal of Personality and Social Psychology: personality Processes and Individual Differences section (http://www.apa.org/journals/psp.html), submit manuscripts to Charles S. Carver, PhD, Department of Psychology, University of Miami, P.O. Box 248185, Coral Gables, FL 33124-2070.

Electronic submission: As of January 1, 2003, authors will be expected to submit manuscripts electronically through the journal’s Manuscript Submission Portal (see the Web site listed above with each journal title). Authors who are unable to do so should correspond with the editor’s office about alternatives.

Manuscript submission patterns make the precise date of completion of the 2003 volumes uncertain. Current editors Leah L. Light, PhD, Stephen N. Haynes, PhD, Ross D. Parke, PhD, Mark E. Bouton, PhD, and Ed Diener, PhD, respectively, will receive and consider manuscripts through December 31, 2002. Should 2003 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2004 volumes.