ADHD and Achievement: Meta-Analysis of the Child, Adolescent, and Adult Literatures and a Concomitant Study With College Students

Thomas W. Frazier, Eric A. Youngstrom, Joseph J. Glutting, and Marley W. Watkins

Abstract

This article presents results from two interrelated studies. The first study conducted a meta-analysis of the published literature since 1990 to determine the magnitude of achievement problems associated with attention-deficit/hyperactivity disorder (ADHD). Effect sizes were significantly different between participants with and without ADHD (sample weighted $r = .32$, sample weighted $d = .71$; $p = .001$). Effects were also examined according to the moderators of age, gender, achievement domain (reading, math, spelling), measurement method (standardized tests vs. grades, parent/teacher ratings, etc.), sample type (clinical vs. nonclinical), and system used to identify ADHD ($DSM-III-R$ vs. $DSM-IV$). Significant differences emerged from the moderator comparisons. The second study, using averaged effect sizes from the first study as a baseline for comparison, investigated achievement levels for an understudied age group with ADHD, namely, college students. Unlike previous studies at the college level, the sample incorporated both student and parent ratings ($N = 380$ dyads). The results were comparable to outcomes from the meta-analysis for college students and adults. Analyses demonstrated modest ($R = .21$) but meaningful predictive validity across 1 year to end-of-first-year grades. However, unlike earlier studies with children and adolescents, student ratings were as predictive as parent ratings. Findings are discussed in terms of the impact of moderator variables on ADHD and achievement.

Poor academic performance is among the most prominent features associated with attention-deficit/hyperactivity disorder (ADHD). Findings reported in a substantial literature have indicated that individuals with ADHD are at risk for a range of academic complications, including a higher incidence of failing grades, elevated rates of grade retention (Fergusson & Horwood, 1995; Fergusson, Lynskey, & Horwood, 1997), an increased occurrence of learning disabilities (LD), and lower scores on standardized tests of achievement (Abikoff & Horwood, 1995; Abikoff, Courneye, Szeibeil, & Koplewicz, 1996; Carlson & Tamm, 2000; Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995; Frankenberger & Cannon, 1999; Gaub & Carlson, 1997; Halperin et al., 1993; Hoza, Pelham, Dobbs, Owens, & Pillow, 2002; Lahey et al., 1998; Purvis & Tannock, 1997, 2000; Seidman, Biederman, Stephan, et al., 1997; Semrud-Clikeman, Guy, Griffin, & Hynd, 2000; Semrud-Clikeman, Steingard, et al., 2000; Tannock, Martinussen, & Frijters, 2000; Zametkin, Liebennauer, & Fitzgerald, 1993). The association with academic problems appears to be specific to ADHD-related behaviors and is not necessarily explained by comorbid conduct disorders (DuPaul et al., 2004; Frick et al., 1991; Hinshaw, 1992; Rapport, Scanlan, & Denney, 1999).

There are five problems in making generalizations about the achievement level of people with ADHD. First, although a substantial research base has been established in the past 40 years, there has been no systematic attempt to integrate results quantitatively across studies that examined both ADHD and achievement. Second, not all individuals with ADHD experience academic deficits. In fact, some investigations have reported higher achievement for individuals with ADHD when compared to controls or to the population expectancy (e.g., Abikoff & Gittelman, 1985; Forness, Youpa, Hanna, Cantwell, & Swanson, 1992; Goldstein, 1987; Sandson, Bachna, & Morin, 2000). Therefore, these positive effects need to be integrated and averaged along with negative outcomes reported elsewhere to provide a more balanced and accurate perspective. Third, although it has been examined previously whether the effects of ADHD are disparate or equivalent across academic content domains (e.g., reading, mathematics, spelling; see Frazier, Demarree, & Youngstrom, 2004), available information is still inadequate. Fourth, the impact of demography (e.g., age and gender effects) as related to both ADHD and achievement is not well understood. Fifth, a range of
other methodological issues need to be considered that could serve to moderate outcomes (e.g., evaluation of attainment and progress through the use of results from standardized tests vs. parent/teacher rating scales or the use of nominal measures of scholastic performance, such as grade retention rates).

With respect to age, academic problems appear to extend beyond childhood, being also of consequence for adolescents, college students, and adults who either were first diagnosed with ADHD during childhood or for whom symptoms appeared later (Faraone, 1996; Frazier et al., 2004). For example, of 6- to 12-year-old children with ADHD who were followed for 10 to 25 years, nearly two thirds continued to manifest at least one of the disabling symptoms of ADHD (i.e., inattention, hyperactivity, impulsivity) as adolescents and adults (Gittelman, Mannuzza, Shenker, & Bonagura, 1985; G. Weiss & Hechtman, 1993). More pertinent here, adults with ADHD had obtained less formal education or lower grades while in high school (Barkley, Fischer, Smallish, & Fletcher, 2002; Mannuzza, Gittelman-Klein, Bessler, Malloy, & LaPadula, 1993).

This article presents results from two interrelated investigations. The first study evaluates the published literature and attempts to determine the presence, direction, and magnitude of achievement effects for individuals with ADHD. More specifically, the first study uses quantitative, meta-analytic procedures to interpret achievement results in terms of average effect sizes, rather than simply according to the presence or absence of significant results or the qualitative pattern of significance levels. This meta-analysis also affords an opportunity to identify how average effect sizes are moderated by the participants’ age, gender, the specific academic domain under consideration, and other variables of interest. The second study takes averaged effect sizes from the first study and employs them as a baseline for comparison to investigate achievement levels for college students, who are one of the most understudied age groups with ADHD.

STUDY 1

Method

Locating Studies

Using the key terms ADHD, ADD, attention deficit, attention-deficit/hyperactivity disorder, and hyperactivity along with each of the linking terms of achievement, reading, math, spelling, language, grades, and education, several procedures were used separately to find as many empirical studies of ADHD and achievement as possible. First, employing the aforementioned locators as well as the names of prominent investigators in the field (Barkley, Biederman, etc.), computer searches were made of the PsycINFO and MEDLINE bibliographic databases. Second, the primary sources identified in this way were examined for other references to appropriate sources.

The underlying rationale of the meta-analysis was to review the most recent literature on ADHD while simultaneously maintaining a sufficient number of studies that results would be representative and stable. The original intention was to cover studies published after 1995, spanning the years coinciding with the publication of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV; American Psychiatric Association, 1994). However, as the investigation proceeded, the timeline was broadened to increase the number of studies and provide more stable results. Therefore, only articles published during the last 15 years were reviewed (i.e., 1990 or after).

Inclusion Criteria

The search was limited to journal articles. Dissertations, book chapters, technical reports, and master’s theses were excluded. Thus, the analysis included only peer-reviewed, empirical investigations as a means to ensure overall quality control (see Reid, Gonzalez, Nordness, Trout, & Epstein, 2004; Weisz, Weiss, Han, Granger, & Morton, 1995). The major decision rule for incorporating a study was that the investigation included individuals with ADHD and reported one or more achievement variables.

Concerning the target samples, participants were identified as having ADHD by one of the following methods:

1. criteria from the Diagnostic and Statistical Manual of Mental Disorders, third revised edition (DSM-III-R; American Psychiatric Association, 1987) or fourth edition (DSM-IV; American Psychiatric Association, 1994);
2. psychiatric/clinical diagnosis;
3. school assessments;
4. performance in the clinical range of ADHD symptomatology, as evaluated by the results of a behavior rating scale; or
5. currently being served in programs for individuals with ADHD.

Nearly all studies (93.5%) either included a typical control group or employed correlations to achievement criteria that obviated the need for controls. When no control group was present, and mean scores were reported (which occurred mostly in studies with adults), a control group was developed, and its mean was set to the population value (M = 100, SD = 15, using Wechsler metric; or M = 0.00, SD = 1.0, using z-score metric). Similarly, because these studies (k = 5) attempted to compare participants to population expectations, sample sizes for the control groups were imputed to 100 to approximate the population estimate (i.e., N) found at most age levels for individually administered tests of achievement (e.g., Anastasi & Urbina, 1997; Kamphaus, 2001; Sattler, 2001; Snelbaker, Wilkinson, Robertson, & Glutting, 2001).

Inclusion criteria went beyond standardized achievement test scores
and incorporated related variables that functioned as proxies for achievement (i.e., grade point average, class ranking, parent/teacher ratings of achievement, failing a grade, provision of extra tutoring, receiving special education services, obtaining a diagnosis of LD, dropping out of school, placement on academic probation while attending college, and the number of semester credits passed in college). Insufficient information was available from the majority of studies to code whether participants had received or were currently receiving treatment (e.g., medication, psychotherapy, or combinations of the two). Likewise, too few studies disaggregated results according to subtypes identified by the DSM-IV (American Psychiatric Association, 1994) to compare each subtype in the meta-analysis.

**Exclusion Criteria**

Unlike school dropout, truancy and suspension or expulsion were not included, because these variables extend beyond academic competence. Furthermore, the meta-analysis did not examine preschool children or studies published prior to 1990.

The initial search yielded 109 investigations with 434 effect sizes (M = 3.98 effect sizes per study). Based on common meta-analytic guidelines (Cooper & Hedges, 1996; Lipsey & Wilson, 2001), several decisions were then made to remove statistical dependencies. Overlap among samples or effect sizes was the greatest threat to statistical independence. Effect sizes in a meta-analysis are assumed to be independent if no more than one effect size is reported for a given sample (Lipsey & Wilson, 2001).

Most studies in the current analysis reported several effect sizes. Therefore, in situations with multiple outcomes, an overall value was obtained by averaging across measures (e.g., reading, mathematics, spelling test scores) to yield a single effect size, which maintained each study’s independence. The modal example of statistical dependence took place when a sample was followed longitudinally. In such instances, effect sizes were usually computed only for the first wave of data collection. When multiple publications were available for the first wave, studies were selected that provided the greatest amount of data (i.e., the largest sample sizes, or roughly equivalent sample sizes with more achievement measures). In several instances, researchers reported results from standardized achievement tests in one investigation (e.g., reading, mathematics, and spelling scores from the Wide Range Achievement Test, third edition; WRAT-3; Wilkinson, 1993), and other achievement-related data (e.g., results from parent rating scales, being retained) in another study. When this happened, both studies were included, but only nonredundant data were retained. The exception to the first-wave rule occurred for age comparisons. When the same sample was evaluated at two or more age levels (e.g., childhood and adolescence), the data were used once and coded into the age group with the smallest number of studies. In the end, every attempt was made to ensure that each study represented an independent sample. The codebook for removing redundancies is available from the senior author.

**Evaluation of Methodological Factors**

Investigations were coded according to several domains of interest to assess whether the comparisons were statistically significant. Specifically, the influence of methodological factors was evaluated by categorizing each study according to the following variables: (a) age of participants (children, adolescents, college students, adults); (b) gender (mixed, male only, female only); (c) achievement area or type of measure (standardized reading, mathematics, and spelling achievement; parent/teacher rating scales; other interval data, such as grade point averages; and categorical outcomes, such as being retained); (d) sample type (clinical vs. nonclinical); and (e) the DSM system employed (indeterminate/nonclinical, DSM-III-R, DSM-IV). Effect sizes were then compared and contrasted among moderator groups using z tests (Lipsey & Wilson, 2001; Wolf, 1986). Likewise, a z test was used to evaluate whether the overall effect size was greater than zero. Coding for all studies was performed by the senior author. The second author then coded 20 studies randomly selected on all methodological characteristics to evaluate reliability. Agreement was 100% for these comparisons.

**Calculation of Effect Sizes**

For each measure, we computed J. Cohen’s (1988) standardized mean difference (d) along with the pooled correlation coefficient (r). Both effect sizes were calculated using formulas provided in Lipsey and Wilson (2001). ADHD participants were recorded as the control group, so that effect sizes (d and r) would remain positive when individuals with ADHD scored lower on the achievement variable. For categorical variables, Cohen’s d was calculated by coding both variables dichotomously (e.g., ADHD = 0, Control = 1; failing a grade = 0, not failing a grade = 1) and calculating the mean for the ADHD and control groups. This is mathematically equivalent to computing a chi square with 1 degree of freedom and subsequently converting chi square to r. Homogeneity analyses were also performed for each measure. A homogeneity analysis is analogous to an analysis of variance, in that homogeneity is the sum of squares of effect sizes about their weighted mean. The homogeneity statistic has a chi-square distribution with k−1 df, where k is the number of contributing effect sizes. Thus, homogeneity evaluates whether the observed effect sizes are likely to result from the sampling of one population. A statistically significant chi square suggests that the effects are not homogenous and that
they may come from more than one distribution.

**Results**

Of the original 109 studies, 72 fulfilled all criteria for inclusion. Multiple investigations \(k = 37\) were excluded because of redundancies, representing 34% of the original database. Table 1 presents distinguishing features of the retained research studies. Of the included samples, 54 studies involved children, 7 evaluated adolescents, 4 examined college students, and 7 compared adults. Insufficient samples were available to analyze the results separately for college students and adults. Consequently, the two age groups were collapsed into a single group labeled adults \(k = 11\) studies. As Table 1 shows, the majority of studies were mixed with respect to gender \(k = 65\), and only four presented results separately for male participants, and another three did so for female participants. Table 1 reveals that most investigations involved clinical samples \(k = 66\) rather than large heterogeneous or epidemiological samples \(k = 6\). Table 1 also shows that the preponderance of studies employed either the DSM-III-R \(k = 33\) or DSM-IV \(k = 18\) criteria to identify individuals with ADHD.

The 72 studies yielded 181 nonredundant effect sizes, where only one averaged reading measure was coded per study, only one averaged mathematics measure was coded, and so on \((M = 2.51\) nonredundant effect sizes per study). The nonredundant effect sizes were themselves then averaged to produce one overall effect per study. Table 1 shows the weighted effect sizes (overall \(r\) and overall \(d\)). Interesting enough, for 3 of the 72 investigations (Hechtman et al., 2004; Sandson, Bachna, & Morin, 2000; Spencer et al., 1995), the overall effect was contrary to expectations and actually showed higher achievement for individuals with ADHD.

Table 2 presents weighted mean effect sizes, confidence intervals, and homogeneity statistics for the overall analysis, separately according to moderator groups. To determine whether individuals with ADHD generally scored lower on measures of achievement, we computed an overall weighted mean effect size \(d\) and used the \(z\) test to evaluate the statistical significance of this effect (Lipsey & Wilson, 2001; Wolf, 1986). Individuals with ADHD displayed significantly lower overall levels of achievement relative to controls, \(d = .71, z = 12.28, p = .001\). According to J. Cohen (1988), a \(d\) of .50 represents a medium effect size, and .80 is the benchmark for a large effect size. Figure 1 presents a stem-and-leaf plot of the distribution of effect sizes for overall achievement.

Rosenthal (1991) recommended a method to test the robustness of effect sizes against sampling bias introduced by the tendency to selectively publish positive results and leave negative results in a file drawer. This method involves calculating the number of unpublished studies with null results that would be necessary to reduce the observed effect size to a level where its meaning changes. In the current case, the weighted mean effect size is unlikely to be a result of publication bias. Rosenthal’s (1991) file-drawer analysis showed that 98 consecutive comparisons with effect size \(d = 0.00\) would be needed to bring the obtained overall effect size \(d = .71\) down to \(d = .30\)—a small effect—and more than 158 consecutive studies with null results would be needed to reduce the mean effect size to \(d = .05\), a negligible effect (Orwin, 1983). Finally, the overall homogeneity statistic was significant, \(\chi^2(71, N = 72) = 446.48, p = .001\), denoting the possibility of moderator variables.

Table 3 presents data on the number of effect sizes per moderator \((k\)\), sample sizes contributing to each effect \((N)\), and weighted mean effect sizes, \(d\). Furthermore, the table shows each \(d\) converted to the familiar Wechsler metric \((M = 100, SD = 15)\). With respect to age comparisons, the table reveals that children showed a significantly larger effect size \(d = .75\) than adolescents \(d = .60, p = .001\), who in turn showed a significantly larger effect size than adults \(d = .57, p = .001\). Thus, children with ADHD generally obtained lower achievement than adolescents, who obtained lower achievement than adults. No significant difference was present for gender. However, an inference of equality is tenuous, because very few studies included either just male \((k = 4)\) or just female participants \((k = 3)\).

The greatest disparities took place according to the achievement domain or assessment methodology. The largest effect occurred in the content area of reading \(d = .73\), followed by mathematics \(d = .67\), and then by spelling \(d = .55\); all three differences were statistically significant at \(p = .001\). Effect sizes were influenced—to an extent—by the type of assessment methodology. Specifically, outcomes evaluated using teacher/parent rating scales resulted in a significantly larger effect \((weighted d = .64)\) than outcomes obtained using either other/interval measures \((e.g., grade point averages, years of education) d = .56, p = .001\, and nominal measures \((e.g., being retained, receiving special education services) d = .49, p = .001\). At the same time, effect sizes were larger for both reading \((d = .73)\) and mathematics outcomes evaluated through standardized tests \((d = .67)\) than they were for achievement outcomes evaluated through rating scales \((d = .64, all p ≤ .001)\).

The type of sample also affected findings. A significantly larger effect was found for large, heterogeneous, or epidemiological samples \((d = .78)\) than when studies concentrated on clinical samples \((d = .68, p = .001\). Finally, a significantly larger effect size was obtained for studies using the DSM-III-R \((d = .79)\) than when the current DSM-IV system was employed to identify individuals with ADHD \((d = .64, p = .001)\.

Perhaps the most striking feature of Table 3 is the uniformity of outcomes. As the table shows, when effect
### TABLE 1
Demographics, Selected Methodological Characteristics, and Effects Sizes of All Studies Included in the Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Age</th>
<th>Gender</th>
<th>Sample</th>
<th>Criteria</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>August &amp; Garfinkel, 1993</td>
<td>65</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III</td>
<td>0.25</td>
</tr>
<tr>
<td>August et al., 1996</td>
<td>111</td>
<td>Ao</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.05</td>
</tr>
<tr>
<td>Barkley, DuPaul, &amp; McMurray, 1990</td>
<td>79</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III</td>
<td>0.41</td>
</tr>
<tr>
<td>Barkley, Fischer, et al., 1990</td>
<td>189</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.33</td>
</tr>
<tr>
<td>Barkley et al., 1991</td>
<td>189</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.49</td>
</tr>
<tr>
<td>Barkley et al., 2002</td>
<td>95</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.44</td>
</tr>
<tr>
<td>Barry et al., 2002</td>
<td>66</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.45</td>
</tr>
<tr>
<td>Benedetto &amp; Tannock, 1999</td>
<td>30</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.35</td>
</tr>
<tr>
<td>Biederman et al., 1999</td>
<td>262</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.30</td>
</tr>
<tr>
<td>Biederman, Faroone, Milberger, Curtis, et al., 1996</td>
<td>218</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.32</td>
</tr>
<tr>
<td>Biederman, Faroone, Milberger, Guite, et al., 1999</td>
<td>237</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.43</td>
</tr>
<tr>
<td>Biederman et al., 1993</td>
<td>291</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.23</td>
</tr>
<tr>
<td>Biederman et al., 1998</td>
<td>120</td>
<td>Ao</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.34</td>
</tr>
<tr>
<td>Biederman et al., 1995</td>
<td>140</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.15</td>
</tr>
<tr>
<td>Bonafina et al., 2000</td>
<td>174</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.32</td>
</tr>
<tr>
<td>Brock &amp; Knapp, 1996</td>
<td>42</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.05</td>
</tr>
<tr>
<td>Casey et al., 1996</td>
<td>162</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.32</td>
</tr>
<tr>
<td>N. J. Cohen et al., 2000</td>
<td>153</td>
<td>Ao</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.26</td>
</tr>
<tr>
<td>Danckaerts et al., 1999</td>
<td>98</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.39</td>
</tr>
<tr>
<td>Dewey et al., 2003</td>
<td>150</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.51</td>
</tr>
<tr>
<td>Faraone et al., 1996</td>
<td>298</td>
<td>Ao</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.18</td>
</tr>
<tr>
<td>Faraone et al., 2002</td>
<td>472</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.19</td>
</tr>
<tr>
<td>Faraone et al., 1998</td>
<td>235</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.44</td>
</tr>
<tr>
<td>Fischer et al., 1990</td>
<td>86</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.47</td>
</tr>
<tr>
<td>Fischer et al., 1993a</td>
<td>169</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.67</td>
</tr>
<tr>
<td>Fischer et al., 1993b</td>
<td>123</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.46</td>
</tr>
<tr>
<td>Frick et al., 1991</td>
<td>265</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.23</td>
</tr>
<tr>
<td>Glutting et al., 2002</td>
<td>680</td>
<td>Co</td>
<td>MF</td>
<td>NC</td>
<td>DSM-IV</td>
<td>0.26</td>
</tr>
<tr>
<td>Greene et al., 1997</td>
<td>150</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.38</td>
</tr>
<tr>
<td>Hechtman et al., 2004</td>
<td>134</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.14</td>
</tr>
<tr>
<td>Heiligenstein et al., 1999</td>
<td>54</td>
<td>Co</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.40</td>
</tr>
<tr>
<td>Jensen et al., 2001</td>
<td>284</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.09</td>
</tr>
<tr>
<td>Krane &amp; Tannock, 2001</td>
<td>169</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.52</td>
</tr>
<tr>
<td>Kroese et al., 2000</td>
<td>44</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.50</td>
</tr>
<tr>
<td>Kuhn et al., 1997</td>
<td>130</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.42</td>
</tr>
<tr>
<td>Lahey et al., 1994</td>
<td>63</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.33</td>
</tr>
<tr>
<td>Lamminmäki et al., 1995</td>
<td>19</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.03</td>
</tr>
<tr>
<td>Latimer et al., 2003</td>
<td>174</td>
<td>Au</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.44</td>
</tr>
<tr>
<td>Livingston et al., 1996</td>
<td>139</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.13</td>
</tr>
<tr>
<td>Mahone et al., 2002</td>
<td>38</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.37</td>
</tr>
<tr>
<td>Mannuzza et al., 1993</td>
<td>158</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III</td>
<td>0.41</td>
</tr>
<tr>
<td>Marks et al., 1999</td>
<td>166</td>
<td>Ao</td>
<td>MO</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.41</td>
</tr>
<tr>
<td>Marshall et al., 1997</td>
<td>122</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.15</td>
</tr>
<tr>
<td>Matochik et al., 1996</td>
<td>121</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.05</td>
</tr>
<tr>
<td>Mayes et al., 2002</td>
<td>94</td>
<td>Co</td>
<td>MF</td>
<td>CL</td>
<td>I/N</td>
<td>0.34</td>
</tr>
<tr>
<td>Mayes et al., 2000</td>
<td>63</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.11</td>
</tr>
<tr>
<td>Merrell &amp; Tymms, 2001</td>
<td>2014</td>
<td>Ch</td>
<td>MF</td>
<td>NC</td>
<td>DSM-IV</td>
<td>0.34</td>
</tr>
<tr>
<td>Mitsis et al., 2000</td>
<td>174</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.30</td>
</tr>
<tr>
<td>Molina et al., 2001</td>
<td>247</td>
<td>Ch</td>
<td>MF</td>
<td>NC</td>
<td>I/N</td>
<td>0.54</td>
</tr>
<tr>
<td>Morgan et al., 1996</td>
<td>130</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.17</td>
</tr>
<tr>
<td>Muir-Broaddus et al., 2002</td>
<td>138</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-IV</td>
<td>0.22</td>
</tr>
<tr>
<td>Nigg et al., 1998</td>
<td>104</td>
<td>Ch</td>
<td>MO</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.32</td>
</tr>
<tr>
<td>Pineda et al., 1999</td>
<td>128</td>
<td>Ch</td>
<td>MF</td>
<td>CL</td>
<td>DSM-III-R</td>
<td>0.68</td>
</tr>
<tr>
<td>Rapport et al., 1999</td>
<td>325</td>
<td>Ch</td>
<td>MF</td>
<td>NC</td>
<td>I/N</td>
<td>0.36</td>
</tr>
</tbody>
</table>
sizes were converted to Wechsler metric, results were homogeneous. For example, sample weighted $d$ values for the DSM-III-R ($d = .79$) versus DSM-IV ($d = .64$) comparison showed a statistically significant difference between systems, $p = .001$. However, when the two $d$s were converted to Wechsler metric, the discrepancy was only 2 points (88 and 90, respectively), suggesting that the differences attributable to the diagnostic criterion, though reliable, are too small to have much impact clinically. In fact, when all the effect sizes in Table 3 were converted to Wechsler metric, results ranged from a low of 88 to a high of 93. This 5-point disparity is within the normal random measurement error (i.e., the standard error of measurement) reported for most individually administered tests of achievement (Anastasi & Urbina, 1997; Kamphaus, 2001; Sattler, 2001; Snelbaker et al., 2001). Therefore, on average, clinicians can expect individuals with ADHD to obtain a standard score of approximately 89 on measures of achievement.

## STUDY 2

Less is known about ADHD at the college level than about ADHD in children, adolescents, and adults (DuPaul et al., 2004; Heiligenstein et al., 1998; Heiligenstein, Guenther, Levy, Savino, & Fulwiler, 1999). Furthermore, there is reason to believe that outcomes obtained with children and adolescents with ADHD may not hold for college students (Glutting et al., 2002, Heiligenstein et al., 1998). College students with ADHD are likely to have (a) higher ability levels, (b) greater academic success during primary and secondary school, and (c) better compensatory skills than individuals with ADHD from the general population. College students with ADHD also experience a different set of stressors than young adults with the condition who do not seek postsecondary training. In particular, college students must adapt to the academic challenges and demands that accompany a college education. Therefore, college students with ADHD may constitute a distinct subset of individuals with the disorder.

The meta-analysis revealed that to date, only four studies have examined the relationships between ADHD and achievement at the college level (Glutting et al., 2002, Heiligenstein et al., 1999; Mayes, 2002; Spinella & Miles, 2003). This lack of inquiry is surprising, because the number of college students with ADHD is growing so fast that their number might soon equal that of students with LD (Latham, 1995). Consequently, our sec-
oord study sought to investigate the 1-year predictive validity of ADHD ratings in forecasting college achievement.

**Method**

**Participants**

The sample comprised dyads (a student and one parent) of first-year students enrolled in degree programs at a university in the northeast corridor of the United States (N = 380). The students came from 18 states. The average age was 19.1 years (SD = 0.37; range = 18–22 years); 127 students (33.4%) were male; 80.5% of the sample was European American, 8.4% African American, 4.7% Hispanic, 4.7% Asian, and 1.7% of other ethnic backgrounds. Although students’ Verbal (M = 567.2, SD = 77.3) and Quantitative (M = 576.4, SD = 78.5) scores from the SAT were slightly above the expectancy for high school students (Verbal scale national M = 520, SD = 110; Quantitative scale national M = 524, SD = 112; College Board, 1995), the scores were representative of averages for entering first-year college students at the university where the study took place. The average high school class rank for students in this study was moderately high (M = 79th percentile, SD = 14.57), and the average grade point average (GPA) at the end of the students first year of college was, as would be expected, moderate (M = 2.87, SD = 0.65). Furthermore, the correlation between high school (HS) class rank and GPA at the end of the first year was significant, r =

---

**Table 2**

Sample Weighted Mean Effect Sizes, Confidence Intervals, and Homogeneity Statistics for Overall Effect Sizes and Effect Sizes Calculated Separately by Moderator Groupings

<table>
<thead>
<tr>
<th>Moderator</th>
<th>k</th>
<th>r</th>
<th>95% CI (r)</th>
<th>d</th>
<th>95% CI (d)</th>
<th>Homogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>72</td>
<td>.32</td>
<td>.31–.33</td>
<td>.71</td>
<td>.70–.72</td>
<td>446.48***</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>54</td>
<td>.30</td>
<td>.31–.32</td>
<td>.75</td>
<td>.74–.76</td>
<td>375.06***</td>
</tr>
<tr>
<td>Adolescents</td>
<td>7</td>
<td>.28</td>
<td>.27–.29</td>
<td>.60</td>
<td>.59–.62</td>
<td>18.05**</td>
</tr>
<tr>
<td>Adults</td>
<td>11</td>
<td>.26</td>
<td>.25–.27</td>
<td>.57</td>
<td>.56–.59</td>
<td>26.02***</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>65</td>
<td>.31</td>
<td>.30–.32</td>
<td>.71</td>
<td>.70–.72</td>
<td>441.66***</td>
</tr>
<tr>
<td>Male only</td>
<td>4</td>
<td>.33</td>
<td>.32–.34</td>
<td>.70</td>
<td>.69–.72</td>
<td>3.12</td>
</tr>
<tr>
<td>Female only</td>
<td>3</td>
<td>.33</td>
<td>.32–.34</td>
<td>.70</td>
<td>.69–.71</td>
<td>1.70</td>
</tr>
<tr>
<td>Achievement area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>71</td>
<td>.31</td>
<td>.20–.32</td>
<td>.73</td>
<td>.72–.75</td>
<td>730.02***</td>
</tr>
<tr>
<td>Mathematics</td>
<td>46</td>
<td>.28</td>
<td>.27–.29</td>
<td>.67</td>
<td>.64–.69</td>
<td>325.65***</td>
</tr>
<tr>
<td>Spelling</td>
<td>19</td>
<td>.25</td>
<td>.24–.26</td>
<td>.55</td>
<td>.53–.57</td>
<td>143.92***</td>
</tr>
<tr>
<td>Other/Interval</td>
<td>18</td>
<td>.27</td>
<td>.26–.28</td>
<td>.56</td>
<td>.56–.58</td>
<td>65.58***</td>
</tr>
<tr>
<td>Rating scales</td>
<td>13</td>
<td>.30</td>
<td>.29–.31</td>
<td>.64</td>
<td>.53–.85</td>
<td>46.18***</td>
</tr>
<tr>
<td>Nominal/Dichotomous</td>
<td>14</td>
<td>.22</td>
<td>.21–.24</td>
<td>.49</td>
<td>.46–.53</td>
<td>148.04***</td>
</tr>
<tr>
<td>Sample type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>66</td>
<td>.29</td>
<td>.28–.30</td>
<td>.68</td>
<td>.67–.69</td>
<td>420.03***</td>
</tr>
<tr>
<td>Nonclinical</td>
<td>6</td>
<td>.34</td>
<td>.33–.35</td>
<td>.78</td>
<td>.77–.79</td>
<td>26.45***</td>
</tr>
<tr>
<td>Criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-III-R</td>
<td>33</td>
<td>.33</td>
<td>.32–.34</td>
<td>.79</td>
<td>.78–.80</td>
<td>256.03***</td>
</tr>
<tr>
<td>DSM-IV</td>
<td>18</td>
<td>.29</td>
<td>.28–.30</td>
<td>.64</td>
<td>.63–.65</td>
<td>83.76***</td>
</tr>
</tbody>
</table>

Note. Most studies reported more than one effect size. In instances where multiple outcomes were provided, effect sizes were averaged across achievement measures (e.g., reading, mathematics, spelling) to yield a single effect that maintained each study’s statistical independence. k = number of effect sizes; CI = confidence interval; r = sample weighted effect sizes presented as correlation coefficients; d = sample weighted effect sizes presented as J. Cohen’s (1988) standardized mean difference; DSM-III-R = Diagnostic and Statistical Manual of Mental Disorders, 3rd revised ed. (American Psychiatric Association, 1987); DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (American Psychiatric Association, 1994). Homogeneity = Rosenthal’s (1991) and Rosenthal and Rubin’s (1982) homogeneity measure, distributed as $\chi^2$ with k – 1 degrees of freedom. Statistical significance suggests that heterogeneity is present and that the coefficients are not likely to come from the sampling of one population of effect sizes.

*Represents variables such as grade point averages, years of education (coded only for studies with adults), high school rank, semester credits passed during either high school or college, and results from standardized tests measuring variables other than reading, mathematics, and spelling (e.g., language, writing, phonemic awareness). Denotes scores obtained from either parent or teacher rating scales. Examples include (a) the Learning Problem scale of the Conners Parent Rating Scales (Conners, 1989); (b) the Academic Achievement Scale of the Personality Inventory for Children–Revised (Witt, Lachar, Klinedinst, & Seal, 1990); and (c) the Learning Problems scale of the Teacher Report Form of the Behavior Assessment System for Children (Reynolds & Kamphaus, 1992). *Represents variables such as (a) the presence or absence of LD; (b) receiving special education services; (c) being retained in school; (d) getting tutoring or extra help; (e) dropping out of school; and (f) passing/failing performance on a high school proficiency test.

**p < .01. ***p < .001.
only 2% of students were diagnosed with ADHD. Similarly, only 2% of students who are at risk for having ADHD. On the other hand, examiners with appropriate training can use the CARE as part of a comprehensive ADHD assessment.

ADHD Measures
Participants and their parents both completed the College ADHD Response Evaluation (CARE; Glutting, Sheslow, & Adams, 2002). The CARE can be used for two purposes, depending on the examiner’s background and training. Its primary use is by postsecondary disability service providers, whose backgrounds may not be in assessment. For these professionals, the CARE can be applied for screening purposes to identify college students who are at risk for having ADHD. On the other hand, examiners with appropriate training can use the CARE as part of a comprehensive ADHD assessment.

The CARE encourages consensual validity because its assessments include conormed student and parent measures: the Student Response Inventory and the Parent Response Inventory. Most often, just the student scale is administered for screening purposes. The results may be interpreted with reference to general national norms for college students, or to gender-specific norms (male vs. female). Each instrument is described hereafter.

Student Response Inventory. The Student Response Inventory (SRI) is a 44-item self-rating scale; 18 items come directly from criteria in the DSM-IV. At the same time, the SRI is relatively brief, taking less than 10 min to complete. Item development was heavily influenced by mental health professionals experienced in working with students with ADHD at the college level. Postsecondary disability service providers, college counselors, psychologists, and psychiatrists with appropriate knowledge and backgrounds were interviewed and asked to write questions.

For each item, students indicate whether they agree, disagree, or are undecided about how an item’s content applies to their day-to-day lives. This format differs from that in some ADHD measures, where symptoms are rated using a 4-point scale (e.g., not at all, just a little, pretty much, very much). The SRI’s use of a neutral or midpoint alternative is consistent with findings that forced-choice systems (e.g., 4-, 6-, or 8-point ratings) result in less response discrimination on personality measures, with raters systematically collapsing the two middlemost alternatives into a single neutral category (Glutting & Oakland, 1993; McKelvie, 1978; Tseng, 1983). Attempts to attain precision by adding a large number of options with a neutral alternative (e.g., 5-point scaling) can also lead to inaccurate responses (McDermott, 1986). This situation arises when raters do not make subtle choices imposed by the item format (e.g., differentiating between gradations such as strongly agree and agree).

Parent Response Inventory. The majority of postsecondary students with ADHD are referred because of difficulties in attention, concentration, and behavioral regulation. These very same problems might also affect their responses to questionnaires. Compounding the problem of response distortion is the finding that students with ADHD underreport key symptoms (Fischer et al., 1993a; Hinshaw, Henker, & Whalen, 1984; Youngstrom, Loeber, & Stouthamer-Loeber, 1999). The net effect is that an ADHD assessment that relies solely on self-report runs certain risks and may under- or overreport clinically important phenomena.

The Parent Response Inventory (PRI) was developed to supplement and enhance data supplied by students on the SRI. The PRI is an objective rating scale completed by a student’s parent. It contains 30 items, 18 of which come directly from ADHD criteria in the DSM-IV. The PRI takes 5 to 10 min to complete and uses the same item format as the SRI, with parents indicating whether they agree, disagree, or are undecided about how an item’s content applies to their child.

Although controversial, the DSM-IV mandates that ADHD symptoms
must be present by age 7 (American Psychiatric Association, 1994). Parents are in a better position to recall the behavior of their offspring as children. Therefore, the PRI asks parents to frame the duration and history of symptoms of their children, using the following retrospective framework: “Please give an opinion about what your son/daughter was like when he/she was in elementary school (approximately 5–8 years old).” The PRI’s use of historical circumstances also helps examiners to clarify whether a student’s difficulties are persistent or whether they are a reaction to stressful events that took place more recently.

**CARE Scales**

The SRI and PRI provide scores based on results from factor analyses as well as scores founded on ADHD criteria in the DSM-IV. Furthermore, each instrument supplies a total score. Thus, the SRI offers six scores (factor-based Inattentiveness, factor-based Hyperactivity, factor-based Impulsivity; DSM-IV Inattentiveness, DSM-IV Hyperactivity; and a Total score), and the PRI supplies five scores (factor-based Inattentiveness, factor-based Hyperactivity/Impulsivity; DSM-IV Inattentiveness, DSM-IV Hyperactivity; and a Total score). Raw scores from the CARE may be transformed to percentiles and T scores (M = 50, SD = 10).

**Reliability and Validity**

The three factor-based scores from the SRI and the two factor-based scores from the PRI were supported through exploratory and confirmatory factor analyses with independent replica-
tions (Glutting, Youngstrom, & Watkins, 2005). The CARE’s manual (Glutting et al., 2002) presents studies examining external validity, including one that evaluated diagnostic validity between college students with ADHD and those without ADHD. Internal consistencies reliabilities for the CARE’s norm sample ranged from .77 to .90 across factor-based scores from the SRI and PRI. Test–retest reliabilities extended from .77 to .88 for the SRI. No test–retest reliabilities were available for the PRI.

**Criterion**

The university supplied binary data on whether students were placed on academic probation at the end of their first year. Specifically, students whose grade point averages (GPAs) were below 2.0 on a 4-point scale were placed on academic probation, whereas students with GPAs greater or equal to 2.0 were not placed on probation.

**Procedure**

Informed consent procedures were approved by the Internal Review Board of the cooperating university. Students and parents were asked in person to complete the SRI and PRI. Ratings were obtained either at the time of new student orientations (at most 2 months prior to the beginning of the fall semester) or when families brought their students to college (the beginning of the fall semester). Respondents were informed that all questionnaires were confidential, but not anonymous, so that it would be possible to track students and obtain information about their academic status. At the request of the school, response rates were not monitored directly. Nevertheless, the participation rate approximated 35% based on the number of returned questionnaires. At the end of the students’ first year, information was obtained about their academic status.

The validity of ADHD ratings was assessed using both bivariate associations and a direct-entry logistic regression analysis (LRA). Only factor-based T scores from the SRI and PRI were employed as predictors. DSM-IV scores were excluded because they overlapped with the factor-based scores and resulted in multicollinearities when both were employed as predictors in the LRA.

**Results**

Table 4 presents bivariate correlations between CARE scores and the GPA criterion of academic probation. Coefficients ranged from .04 to .17. Interestingly, both student-rated Inattentiveness and parent-rated Inattentiveness correlated .17 with probation status. The equal validity found here runs counter to previous studies, in which parent ratings were generally more accurate and predictive (Achenbach et al., 1987; Bird, Gould, & Staghezza, 1992; Loeb, Green, Lahey, & Stouthamer-Loeber, 1989).

Logistic regression, unlike simple bivariate correlations, more accurately illuminates the full network of variable interrelationships that can occur among a group of predictors and a criterion (Tabachnick & Fidell, 2001). A test of the complete model, with all five predictors, against a constant-only model was statistically significant, \( \chi^2(5, N = 327) = 14.16, p = .015 \), indicating that the predictors, as a set, distinguished college students on academic probation from those with average to above-average achievement. Table 5 shows standardized regression coefficients, significance levels, and odds ratios for each predictor. As with the bivariate comparisons, two variables made significant contributions to the prediction of academic status. These variables were student-rated Inattentiveness \((p = .02; \beta = .040)\) and parent-rated Inattentiveness \((p = .05; \beta = .036)\), and their contribution remained significant after controlling for covariation among the other predictors and criterion.

At first glance, the overall prediction seems modest \((R = .211, R^2 = .045)\). However, it is comparable to the results from the meta-analysis, where the association between ADHD and achievement was \(r = .26\) for adults (see Table 2). Furthermore, the current criterion is binary and, therefore, is likely to attenuate associations compared to a continuous measure of achievement (Allen & Yen, 1979).

To further examine the predictive validity of student and parent reports, we computed regressions similar to

---

**TABLE 4**

Summary of Bivariate Correlations Between CARE Scale Ratings and First-year Grade Point Average

<table>
<thead>
<tr>
<th>Predictor</th>
<th>( r )</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Rating Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>.17</td>
<td>.029</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.04</td>
<td>.002</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>.05</td>
<td>.003</td>
</tr>
<tr>
<td>Parent Rating Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>.17</td>
<td>.029</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.07</td>
<td>.005</td>
</tr>
</tbody>
</table>

Note. CARE = College ADHD Response Evaluation (Glutting, Sheslow, & Adams, 2002).

**TABLE 5**

Summary of Logistic Regression Analysis of CARE Scale Ratings to First-year Grade Point Average

<table>
<thead>
<tr>
<th>Predictor</th>
<th>( \beta )</th>
<th>( p )</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Rating Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>.040</td>
<td>.02</td>
<td>1.041</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.005</td>
<td>.80</td>
<td>1.006</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>.016</td>
<td>.46</td>
<td>1.009</td>
</tr>
<tr>
<td>Parent Rating Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inattentiveness</td>
<td>.036</td>
<td>.05</td>
<td>1.037</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.005</td>
<td>.81</td>
<td>1.005</td>
</tr>
</tbody>
</table>

Note. CARE = College ADHD Response Evaluation (Glutting, Sheslow, & Adams, 2002).
those described earlier, except that the dependent variables were SAT total scores and high school rank in separate analyses. For SAT total scores, the results indicated that the predictors, as a set, made a significant prediction, $F(5, 374) = 2.71, p = .020, R = .19$. However, the only significant unique predictor was student-rated Inattentiveness ($p = .027, \beta = .146$) after controlling for the other predictors. Student-rated Hyperactivity ($p = .052, \beta = .127$) and Impulsivity ($p = .054, \beta = .125$) approached significance. For high school rankings, the results indicated that the predictors, as a set, made a significant prediction, $F(5, 321) = 2.37, p = .040, R = .19$. However, the only significant unique predictor was parent-rated Inattentiveness ($p = .018, \beta = .154$). These findings further support the importance of ratings of inattentiveness in the prediction of academic performance.

A better way to evaluate the practical significance of findings is through the interpretation of odds ratios and a case study. Table 5 shows that the odds ratio was 1.041 for student-rated Inattentiveness. Assume there is a student who obtained a T score of 70 on the SRI Inattentiveness scale at the beginning of the school year. This score is 2 SD above the mean. In such a scenario, the student is 34% more likely to end up on academic probation (i.e., have a GPA less than 2.0 on a 4-point scale) than a comparable student without self-reported inattention. This case study demonstrates that ADHD variables meaningfully predicted end-of-year grades. More important, these outcomes suggest that postsecondary counselors and clinicians should routinely screen students for ADHD symptomatology.

**GENERAL DISCUSSION**

The present studies produced several important findings. Foremost, the meta-analytic results indicated a moderate to large discrepancy in academic achievement between individuals with ADHD and typical controls (weighted $d = .71$). This outcome substantiates the significant impact of ADHD symptoms on academic performance, and it reveals a pattern of impairments beyond the achievement test decrements identified previously (Frazier et al., 2004). However, it should be noted that this effect size may overestimate the actual effect size, as unpublished work was not included in this weighted average. It is also possible that some of the individuals included in the studies reviewed may not have met all of the criteria for the DSM-IV ADHD diagnosis. With these caveats in mind, the findings point to significant moderating effects among demographic and methodological variables. Each effect provides important information regarding the impact of ADHD on academic achievement and will be discussed in turn.

Standardized achievement tests—particularly reading measures—produced the largest effect sizes. The obtained pattern of greater deficits in reading is the opposite of that observed in a recent meta-analysis of neuropsychological test performance in ADHD (Frazier et al., 2004). The difference between these meta-analyses may be explained by the fact that the present study did not attempt to exclude individuals with comorbid LD, whereas the previous study did so. Alternatively, the current results may suggest that ADHD symptoms have a greater negative impact on reading performance. The latter interpretation is consistent with the idea that math and spelling generally require the checking of an individual’s work as part of the typical process, whereas many reading tasks can be performed more automatically (i.e., without checking of work) or by obtaining the gist of material. If true, the obtained outcomes suggest that individuals with ADHD will detect mistakes more frequently during math and spelling tasks than during reading. Greater impairment on reading measures may be also due to the significant association between reading difficulties and ADHD (Fergusson & Horwood, 1992). Alternatively, greater impairment on reading measures may have been strictly due to the fact that the present study did not attempt to exclude participants with LD, many of whom had difficulties with reading. Future research is needed to examine these possibilities and to determine the magnitude of ADHD influences on achievement independent of specific learning difficulties.

Co-occurring factors may account for the finding of greater deficits on standardized tests. Standardized achievement tests may be sensitive to both (a) the general effects of ADHD symptoms on everyday learning and retention, and (b) the specific effects of ADHD symptoms on test performance (Glutting, Youngstrom, Oakland, & Watkins, 1996). For example, students might perform poorly on a standardized mathematics test simply as a consequence of errors in attention that also are present in their day-to-day activities (e.g., rounding mistakes or failing to carry integers on simple math problems). Clinically, the second possibility is apparent during evaluations of college students who sometimes obtain low scores on standardized achievement tests because of ADHD but then show high performance in specialized academic courses.

Rating scales and objective interval measures (e.g., GPA, years of education) also produced moderate to large effect sizes; nominal scales (e.g., dropping out of school, repeating a grade) produced smaller, but still medium-sized effects. These effects, in combination, suggest that weaknesses in academic performance generalize beyond standardized tests. Thus, the lower scores on achievement tests obtained by individuals with ADHD cannot be completely attributed to situational factors (e.g., negative test-taking behavior) and are likely to represent real deficits. The consistency of effects across rating scales and objective measures (e.g., standardized tests and interval scales such as GPA) indicates similar sensitivity to academic deficits in ADHD. The discrepancy between nominal scales and other academic measures may be accounted for by the
attenuation of effect sizes due to a restriction of range in the nominal variables. Therefore, future research should not employ nominal indices as the sole achievement measure, because they may underestimate the effects of ADHD. It should also be noted that specific achievement domain effect sizes may overestimate the true effect size, as the present meta-analysis did not include unpublished studies, the inclusion of which might have slightly attenuated weighted mean effect sizes. However, the file-drawer analysis suggests that if such attenuation did influence estimates, this influence is likely to have been relatively minor.

Moderator analyses revealed a trend toward decreasing academic impairment with age. This finding contradicts a cumulative-deficit hypothesis. Three possibilities may account for the disparity. First, there were fewer older samples in the meta-analysis, and they may have had less impairments as a consequence of unmeasured differences, perhaps in selection processes or in the attrition of students with ADHD due to academic failure. The students with the highest impairment by ADHD in terms of achievement are the least likely to persist through high school or to pursue further education afterward. Second, earlier studies have documented a decrease in ADHD symptomatology with age (for a review, see Barkley, 1998), which, in turn, may result in fewer achievement problems among adolescents and adults. Third, there may be greater compensation for the effects of ADHD with age. For example, individuals with ADHD may “learn” about their problems and compensate by checking their work, rereading passages, or seeking additional instruction.

Studies employing the DSM-III-R as the standard for ADHD showed a significantly larger effect size than those using the DSM-IV. It may be the case that the DSM-IV’s inclusion of multiple subtypes resulted in individuals with less impairment meeting the criteria for ADHD. Such an inference is supported by the fact that the DSM-III-R required 8 of 14 ADHD symptoms, whereas the DSM-IV necessitates only 6 of 9 inattentive or 6 of 9 hyperactive/impulsive symptoms.

All of the aforementioned moderating effects produced differences that were statistically significant. Nevertheless, the magnitude of the differences was small. For instance, when effect sizes in the meta-analysis were converted to a Wechsler metric (M = 100, SD = 15), the results across all demographic and methodological categories ranged from a standard score low of 88 to a high of 93. This 5-point disparity is within the normal random measurement error (i.e., the standard error of measurement) reported for most individually administered tests of achievement (Anastasi & Urbina, 1997; Kamphaus, 2001; Sattler, 2001; Snellbaker et al., 2001).

The meta-analysis revealed significant heterogeneity of effects within most moderator groups. This finding is not surprising due to the collapsing of multiple measures in each category. For example, several different achievement tests were employed across studies in the reading, math, and spelling categories. Similarly, multiple measures were collapsed in the subjective rating scale category. Also, the observed heterogeneities may be accounted for by other, unmeasured characteristics, such as methodological quality and idiosyncratic sample characteristics. Future meta-analytic studies and original research may help discern how such differences affect both ADHD and achievement.

The second study also produced important findings by evaluating the effects of both student- and parent-reported ADHD symptoms on achievement with a relatively understudied population, namely, college students. To our knowledge, no prior study has examined external validity for both student and parent ratings at the college level. The results indicated that ADHD symptoms continue to be significantly associated with problems in academic functioning at the college level. The magnitude of the effect (R = .21) was similar to that observed in other studies identified in the meta-analysis (average r = .26 for adults). This difference may be accounted for by two factors: Both the dichotomization of the dependent variable (GPA) and the overrepresentation of female students are likely to have attenuated the true relationship. Therefore, the observed r is not likely to indicate any true difference in the magnitude of effects of ADHD on academic achievement between the present and previous studies. Thus, ADHD symptoms continue to influence academic performance in a relatively selective sample of college students. The current findings also indicate the potential value in the routine screening of college students for ADHD to circumvent the academic failure associated with the disorder.

Interesting enough, both student- and parent-rated inattentiveness correlated equally with probation status (r = .17). This finding of equal validity runs counter to previous studies, in which parent ratings were generally more accurate and predictive (Achenbach et al., 1987; Bird et al., 1992; Loeb, Green, Lahey, & Stouthamer-Loeb, 1989). The equality found here may be a consequence of the age of informants or of their ability level. The current investigation employed self-ratings from college students, whereas earlier studies examined self-ratings from children and adolescents. Moreover, the relative predictive ability of student and parent reports may vary as a function of the criterion, as we found that student reports were particularly predictive when SAT total scores were examined, and parent reports were more important when high school rank served as the criterion. A third possibility is retrospective distortion. Parents in the present study were required to recall events from the time when their children were between 5 and 8 years old.

In contrast to inattentiveness, hyperactive and impulsive symptoms produced relatively little independent...
information. It may be that the influence of hyperactivity and impulsivity decreases with age. Alternatively, parents and students may be less able to accurately report symptoms of hyperactivity or impulsivity. Future research should evaluate the latter possibility by examining the relationship between subjective and objective measures of inattention, hyperactivity, and impulsivity and academic achievement at the college level. Such studies will assist in determining whether the relative importance of inattentive symptoms is an artifact of the method of measurement or whether these symptoms are the most disabling and should be the primary focus of intervention. Future work should also examine the relationship between attention problems, hyperactivity/impulsivity, and achievement in more discrete subsamples (e.g., ADHD with no LD, ADHD with LD, no ADHD) to better understand whether the relationship between inattentiveness and achievement observed in the present study is primarily due to learning problems or to the comorbidity of ADHD and learning problems. These studies should also evaluate medication status, as this may significantly influence the effects of ADHD on achievement.

ABOUT THE AUTHORS

Thomas W. Frazier, PhD, is an associate staff member at The Cleveland Clinic. His current research interests include cognitive and affective assessment and core illness features in attention-deficit/hyperactivity disorder and pediatric bipolar disorder. Eric A. Youngstrom, PhD, is an associate professor in psychology at the University of North Carolina at Chapel Hill. His current research interests include developmental psychopathology, clinical assessment, and evidence-based medicine, with a particular focus on bipolar disorder across the life cycle. Joseph J. Glutting, PhD, is a professor of research, measurement, evaluation, and school psychology at the University of Delaware. His current research interests include psychoeducational assessment, statistical analysis of time-series data as applied to individual case studies, and the valid assessment of bilingual Hispanics. Marley W. Watkins, PhD, is a professor in the School Psychology Program at the Pennsylvania State University and a diplomate of the American Board of Professional Psychology. His research interests include assessment and professional issues in school psychology. Address: Thomas W. Frazier, Section of Behavioral Medicine, The Cleveland Clinic, Shaker Medical Campus, 2801 Martin Luther King Jr. Drive, Cleveland, OH 44104; e-mail: fraziet2@ccf.org

REFERENCES

Note. Asterisks represent studies included in the final meta-analysis, where \( k = 72 \).


Biederman, J., Faraone, S., Milberger, S,


