Assessment of Attention-Deficit/Hyperactivity Disorder: An Evaluation of Six Published Rating Scales

Lehigh University

Abstract. Two key characteristics of Attention-Deficit/Hyperactivity Disorder (ADHD), problems with attention and behavioral control, represent the most common reasons for school referrals (Barkley & Edwards, 1998). This makes it essential for school psychologists to acquire the knowledge and skills necessary to conduct a comprehensive assessment of ADHD and ADHD-related symptoms. The current article reports a critical analysis of six published behavior rating scales commonly utilized in a best practices approach to a school-based comprehensive assessment of ADHD (DuPaul & Stoner, 1994). Each of the rating scales was evaluated for strengths and limitations with regard to purpose, content, standardization, and psychometric properties. Recommendations are delineated regarding the use of each rating scale with specific target populations (i.e., culturally diverse students) as well as specific stages of assessment within a problem-solving process.

Behavior rating scales represent an efficient method for gathering information about children and youth (Wilson & Reschly, 1996). Information obtained from rating scales can be used for several purposes including screening, diagnosis, or monitoring the effects of treatment. As a result of their utility, rating scales completed by adult informants have become a common method for identifying children with behavior problems including Attention-Deficit/Hyperactivity Disorder (ADHD; Power & Eiraldi, 1998). ADHD is a neurological disorder characterized by a persistent pattern of inattention and/or hyperactivity-impulsivity and occurs in 3 to 7% of school-age children (American Psychiatric Association [APA], 2000). A diagnosis of ADHD is determined by establishing the developmental deviance and pervasiveness of symptoms, level of impairment, age of onset, and ruling out of alternative explanations for deviations in child behavior (APA, 2000). Used within this context, behavior rating scales help establish the severity of ADHD-related behaviors across settings relative to a normative sample of children of the same age and gender.

Pediatricians and psychiatrists historically have assumed the primary role in diagnosis and treatment of ADHD (Barkley & Edwards, 1998). However, two of the most common reasons for school referrals, problems with attention and behavioral control, make it essential for school psychologists to acquire knowledge and skills in the assessment of ADHD symptomatology (Landau & Burcham, 1995). Best practices in the school-based assessment of ADHD recommend that school-based practitioners use an ongoing, problem-solving process that leads to the development of effective interventions (Hoff, Doepke, & Landau, 2002). DuPaul and Stoner (1994) recommend several specific stages of assessment for ADHD, and
rating scales are considered relative to three of these: (a) screening, (b) multimethod assessment/comprehensive evaluation, and (c) progress monitoring or program evaluation.

Considerations in Conducting a School-Based Assessment of ADHD

Selecting an appropriate rating scale can be a confusing task given the extensive array of rating scales currently available. In addition to the purpose or stage of assessment and age of the child, the psychometric integrity of instruments and the adequacy of normative data with respect to the child’s culture are essential considerations when selecting an appropriate instrument.

Purpose of assessment. Information obtained from screening activities, such as interviews and rating scales, helps determine the need for a more comprehensive evaluation (i.e., multimethod assessment). A multimethod assessment is designed to determine the developmental deviance of the child’s behavior as well as pervasiveness of impairment across settings. Strategies used in this stage include behavior ratings, direct observations, interviews, review of school records, and assessment of academic skills (DuPaul & Stoner, 1994). Combining various assessment strategies helps minimize the limitations associated with any one method of measurement. Both broad- and narrow-band measures can be utilized in the first two phases of assessment. Narrow-band measures assess specifically the symptoms of ADHD, whereas broad-band measures assess comorbid features and assist with the process of differential diagnosis (Volpe & DuPaul, 2001).

Finally, progress monitoring, as its name implies, serves as a measurement of change associated with intervention. Assessment in this stage is often brief in nature and is scheduled to occur frequently (i.e., daily or weekly). Thus, sensitivity to change over time is an essential feature for rating scales used in this phase of assessment.

Assessing ADHD with culturally diverse students. The assessment of ADHD with culturally diverse students is receiving increased attention in the professional literature (Reid, 1995). It is vital, at a minimum, that the normative data for a rating scale are representative of the population with which it is to be used (Salvia & Ysseldyke, 1988). The possibility of cross-cultural differences has been demonstrated (e.g., Reid et al., 1998), but it has been difficult to determine if reported differences have been due to actual behavior differences, instrument bias, or a combination of the two (Reid et al., 1998). Ideally, evidence regarding the reliability and validity of an instrument should be available for all intended populations.

The primary purpose of this article is to provide a critical review of six published rating scales commonly used for assessing the behavior of students thought to have ADHD. Information provided in this review is intended to assist school psychologists in selecting the most appropriate rating scale based upon the characteristics of a given child (age, race, broad-band symptom profile) and the type of assessment to be undertaken (e.g., screening, diagnostic assessment, progress monitoring).

Instrument Selection and Criterion for Evaluation

An initial search of electronic databases (e.g., ERIC Assessment Clearinghouse, PsycInfo), Buros Mental Measurements Yearbooks, and recent test catalogs yielded nine instruments to be considered for this review. Behavior rating scales included in this review had to: (a) be identified by the developer as appropriate for the screening and/or diagnosis of ADHD, (b) include a parent and teacher version, and (c) be published by a test publisher. Six of these nine instruments met the inclusion criteria: the ADHD Rating Scale-IV (DuPaul, Power, Anastopoulos, & Reid, 1998); Behavior Assessment System for Children: Monitor for ADHD (BASC Monitor; Kamphaus & Reynolds, 1998); Conners’ Rating Scales-Revised (CRS-R; Conners, 1997); ADHD Symptom Checklist-4 (SC-4; Gadow & Sprafkin, 1997); Attention Deficit Disorders Evaluation Scale (ADD-ES-2nd edition; McCarney, 1995); and ADD-H Comprehensive Teacher’s Rating Scale (ACTeRS; Ullmann, Sleator, & Sprague, 1997).
Three instruments identified in the initial search failed to meet the criteria for inclusion in the review. Specifically, two of the instruments, the Attention-Deficit Hyperactivity Disorder Test (ADHD; Gilliam, 1995) and Spadafore ADHD Rating Scale (S-ADHD-RS; Spadafore & Spadafore, 1997), do not include multiple versions of the scale, and the Attention Deficit Disorder Behavior Rating Scale (ADDBRS; Owens, 1998) is published independently by its author. For the purpose of this review, each rating scale was evaluated with regard to content and use, standardization and normative data, and psychometric properties. Characteristics of each scale, including number of items, completion time, rating format, age range, size of standardization sample, and subscales are presented in Table 1. Reference to the characteristics of each scale is limited in the text to facilitate a parsimonious, yet detailed, review of the psychometric properties. Table 2 highlights the psychometric properties of each scale, and Table 3 provides a summary of the strengths, weaknesses and recommended uses.

ADHD Rating Scale-IV

The Home and School versions of the ADHD Rating Scale-IV (ADHD-IV; DuPaul, Power, Anastopoulos, & Reid, 1998) are brief norm-referenced questionnaires designed to gather information from parents and teachers with regard to child symptoms of ADHD in the past 6 months. The ADHD-IV 6 comprises two 9-item subscales (Inattention, Hyperactivity-Impulsivity) that conform to the DSM-IV subtypes of ADHD, and items on the Home and School versions are identical.

The Home and School forms of the ADHD-IV facilitate the gathering of data from multiple settings, and the brevity of the forms makes them potentially useful for monitoring treatment effects. Unfortunately, there currently is little empirical evidence for the instrument’s treatment sensitivity. The authors, however, do provide statistical guidelines and resources for using the instrument to monitor changes in behavior (DuPaul, Anastopoulos et al., 1998).

Standardization sample and norms. The standardization samples for the Home and School versions of the ADHD-IV each consisted of 2,000 children and adolescents (ages 4–20) drawn from over 20 school districts across the United States. Each sample approximated the 1990 U.S. Census data for race and geographic region. Scores on both versions of the ADHD-IV were found to vary significantly by age, gender, and ethnic group. These findings led the authors of the ADHD-IV to provide separate norms by age and gender. It should be noted that DuPaul, Power, McGoe, Ikeda, & Anastopoulos (1998) found African American children to receive higher ratings of ADHD symptoms than Caucasian and Latino children. Given that other studies have found similar results, one should exercise caution when interpreting ratings on the ADHD-IV for African American children. Specifically, the use of the extant normative data in the assessment of African American children may overestimate the severity of their ADHD symptoms.

Psychometric properties. The stability and internal consistency of the ADHD-IV generally are strong. These findings were based on a study of 71 students (35 males, 36 females) between 5 and 17 years of age. Test-retest coefficients (4-week latency) ranged from between .88 and .90 for the School Version to between .78 and .86 for the Home Version. Coefficient alphas (.86–.92) were consistently high for both versions of the checklist (DuPaul, Power, McGoe et al., 1998).

Based upon exploratory and confirmatory factor analyses, it appears that both a two-factor (e.g., Inattention and Hyperactivity/Impulsivity) and a single-factor solution (ADHD) fit the standardization data well (DuPaul, Power et al., 1997; DuPaul, Anastopoulos et al., 1998). The authors chose to utilize two subscales because some analyses favored the two-factor solution in correspondence with the DSM-IV model.

In terms of predictive validity, the ADHD-IV subscales of Inattention (ADHD-I) and Hyperactivity/Impulsivity (ADHD-II) were found to differentiate significantly between children diagnosed with ADHD-Combined, ADHD-I, and psychiatric controls (DuPaul, Power, McGoe et al., 1998). The
Table 1  
Characteristics of ADHD Rating Scales

<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Number of Items</th>
<th>Completion Time (min)</th>
<th>Rating Format</th>
<th>Age Range</th>
<th>Standardization Sample</th>
<th>Subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD Rating Scale-IV (ADHD-IV; DuPaul et al., 1998)</td>
<td>18 (Both Forms)</td>
<td>10</td>
<td>4-point frequency</td>
<td>5-18</td>
<td>2,000 (Teacher)</td>
<td>Inattention, Hyperactivity/Impulsivity Total</td>
</tr>
<tr>
<td></td>
<td>47 (Teacher)</td>
<td>10-15</td>
<td>4-point frequency</td>
<td>4-18</td>
<td>2,401 (Teacher)</td>
<td>Attention Problems, Hyperactivity, Internalizing Problems, Adaptive Skills</td>
</tr>
<tr>
<td></td>
<td>46 (Parent)</td>
<td></td>
<td></td>
<td></td>
<td>3,483 (Parent)</td>
<td></td>
</tr>
<tr>
<td>Conners' Rating Scales-Revised (CRS-R; Conners, 1997)</td>
<td>Short Forms</td>
<td>10-20</td>
<td>4-point frequency</td>
<td>3-17 (Parent/Teacher)</td>
<td>3,870 (Teacher)</td>
<td>Oppositional, Cognitive/Inattention, Hyperactivity, Anxious-Shy, Social Problems, Perfectionism, Psychosomatic, Global Index, ADHD Index, DSM-IV, Symptom Subscales</td>
</tr>
<tr>
<td></td>
<td>28 (Teacher)</td>
<td></td>
<td></td>
<td></td>
<td>4,908 (Parent)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 (Parent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 (Student)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long Forms</td>
<td></td>
<td></td>
<td></td>
<td>6,880 (Student)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59 (Teacher)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>80 (Parent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>87 (Student)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD Symptom Checklist-4 (SC-4; Gadow et al., 1997)</td>
<td>50 (Both Forms)</td>
<td>10-15</td>
<td>4-point frequency</td>
<td>3-18</td>
<td>2,715 (Teacher)</td>
<td>Inattentive, Hyperactive/Impulsive ODD, Peer Conflict, Stimulant Side Effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,844 (Parent)</td>
<td></td>
</tr>
<tr>
<td>Rating Scale</td>
<td>Number of Items</td>
<td>Completion Time (min)</td>
<td>Rating Format</td>
<td>Age Range</td>
<td>Standardization Sample</td>
<td>Subscales</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>-----------------------</td>
<td>--------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Attention Deficit Disorders Evaluation Scale (ADDES; 2nd ed.; McCarney, 1995)</td>
<td>60 (School)</td>
<td>15-20</td>
<td>5-point frequency</td>
<td>4–18 (School)</td>
<td>5,795 (School)</td>
<td>Inattentive Hyperactive/Impulsive</td>
</tr>
<tr>
<td></td>
<td>46 (Home)</td>
<td></td>
<td></td>
<td>3–18 (Home)</td>
<td>2,415 (Home)</td>
<td></td>
</tr>
<tr>
<td>ADD-H Comprehensive Teacher's Rating Scale (ACTeRS; Ullman et al., 1997)</td>
<td>24 (Teacher)</td>
<td>10</td>
<td>5-point frequency</td>
<td>5–14</td>
<td>2,362 (Both Forms)</td>
<td>Attention Hyperactivity Social Skills Oppositional Behavior Early Childhood (Parent only)</td>
</tr>
<tr>
<td></td>
<td>25 (Parent)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability and Validity Evidence for ADHD Rating Scales</td>
<td>Internal Consistency</td>
<td>Interrater</td>
<td>Convergent/Discriminant</td>
<td>Criterion-Related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reliability Evidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-Retest Stability</td>
<td>.88–.90 (4 weeks)</td>
<td>.78–.90 (4 weeks)</td>
<td>Parent</td>
<td>.72–.93 (2–8 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD-IV Teacher</td>
<td>.88–.96</td>
<td>.86–.92</td>
<td>CFA (Parent-Teacher)</td>
<td>.41–.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASC-M Teacher</td>
<td>.72–.93</td>
<td>.77–.93</td>
<td>CFA (Parent)</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRS-R Teacher</td>
<td>.60–.90</td>
<td>.57–.84</td>
<td>CFA (Parent)</td>
<td>Not reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Validity Evidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Structure</td>
<td>DSM-IV</td>
<td>DSM-IV</td>
<td>CTR-S-R</td>
<td>CTR-S-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergent</td>
<td>EFA and CFA 2-factor model</td>
<td>EFA and CFA 2-factor model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discriminant</td>
<td>EFA and CFA 2-factor model</td>
<td>CFA 4-factor model</td>
<td>CTR-S-R (-.36-.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>DSM-IV</td>
<td>DSM-IV</td>
<td>CBCL</td>
<td>CBCL (-.68-.79)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 2 (continues)*

Group differences between children with ADHD and children with emotional problems.
(Table 2 continued)

<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Reliability Evidence</th>
<th>Validity Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test-Retest Stability</td>
<td>Internal Consistency</td>
</tr>
<tr>
<td>SC-4 Teacher</td>
<td>.70–.89 (6 weeks)</td>
<td>.92–.95</td>
</tr>
<tr>
<td>Parent</td>
<td>.75–.82 (6 weeks)</td>
<td>.93–.95</td>
</tr>
<tr>
<td>ADDES-2nd Teacher</td>
<td>.88–.97 (30 days)</td>
<td>.98–.99</td>
</tr>
<tr>
<td>Parent</td>
<td>.88–.91 (30 days)</td>
<td>.96–.98</td>
</tr>
<tr>
<td>ACTeRS Teacher</td>
<td>Not reported</td>
<td>.92–.97</td>
</tr>
<tr>
<td>Parent</td>
<td>Not reported</td>
<td>.78–.96</td>
</tr>
</tbody>
</table>

*Note. LR = Literature Review, ER = Expert Review, EFA = Exploratory Factor Analysis, CFA = Confirmatory Factor Analysis.*
<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Recommended Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD-IV</td>
<td>Well-validated measure Items closely approximate symptom criteria for ADHD Relatively short (18 items) Large normative samples Spanish translation</td>
<td>No data on treatment sensitivity Ethnic differences were found in examination of normative data, yet scoring is based only on age and gender. Specifically, the ADHD-IV may overestimate the symptomatology of African American children</td>
<td>Appropriate for screening and diagnostic assessment. Use caution with African American children and youth Potentially useful for treatment monitoring, although sensitivity to treatment has not been established</td>
</tr>
<tr>
<td>BASC-M</td>
<td>Useful software that allows practitioners to track behavior change of a number of students (see limitations) Spanish translation</td>
<td>Does not utilize all DSM-IV criteria for ADHD in generating norm-based scores Normative data were generated from the original BASC (a broad-band measure No treatment sensitivity data are available. Not recommended for monitoring treatment effects No validity data available for adolescents</td>
<td>Evidence for screening with ages 6–11, but insufficient evidence for other age ranges Potentially useful for treatment monitoring, although sensitivity to treatment has not been established</td>
</tr>
<tr>
<td>CRS-R</td>
<td>Well-validated measure Parent, teacher, and adolescent versions available Sensitive to treatment effects Spanish translation</td>
<td>Minority children were underrepresented in normative samples Limited evidence for self-report form</td>
<td>The Short Form of the Conners is recommended for monitoring treatment effects if a narrow assessment of behavior problems is appropriate The Long Form is useful for diagnostic assessment and offers broad coverage of behavioral symptoms</td>
</tr>
<tr>
<td>SC-4</td>
<td>Inclusion of DSM-IV ODD symptom items, interpersonal aggression items, and a checklist of stimulant side-effects Demonstrated sensitivity to treatment Spanish version</td>
<td>Manual contains reports of many studies and can be complicated Minority children underrepresented in some age groups No structural validity data</td>
<td>Recommended for monitoring treatment effects—especially stimulant medication</td>
</tr>
</tbody>
</table>

(Table 3 continues)
### (Table 3 continued)

<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Recommended Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDES</td>
<td>Specificity of items and response format eases use for informants Manual is user friendly</td>
<td>Normative data for the Home Version needs to be updated Normative data not analyzed for racial differences Validity data based on studies with small samples</td>
<td>Appropriate for use in the diagnostic/comprehensive phase of assessment—particularly when in need of precise information regarding potential target behavior</td>
</tr>
<tr>
<td>ACTeRS</td>
<td>Relatively short (24–25 items) Allows assessment of social skills and oppositional behavior Several published studies demonstrating treatment sensitivity</td>
<td>No recent normative data for parent informants Specific demographics of normative sample not reported. Cannot determine representativeness of the sample Limited validity data</td>
<td>Not recommended for use at this time</td>
</tr>
</tbody>
</table>

ADHD-IV also appears to correlate well with other rating scales commonly used in the assessment of ADHD. For example, significant positive correlations (.66–.85) have been demonstrated between the Home and School versions of the ADHD-IV and relevant subscales of the Conners Parent Rating Scale-48 (CPRS) and Conners Teacher Rating Scale-39 (CTRS), respectively (DuPaul, Power, McGoey et al., 1998). Significant correlations also have been demonstrated between the Inattention and Total scores of the School Version of the ADHD-IV and several objective measures gathered in the school setting such as behavioral observations and worksheet accuracy (DuPaul, Power, McGoey et al., 1998). The Hyperactivity/Impulsivity score, however, correlated significantly (r = .34) with worksheet accuracy only.

The predictive validity of the ADHD-IV was investigated with a sample of 92 clinically referred children between the ages of 6 and 15 years (Power et al., 1998). The School Version appeared somewhat better than the Home Version in predicting diagnostic status, whether the distinction was being made between ADHD subtypes (e.g., ADHD-I vs. ADHD-CM) or between ADHD groups and a clinical control group (children referred to a clinic who did not meet research diagnostic criteria for any ADHD subtype). Specifically, the percent of cases classified correctly ranged between 65 to 84% for the School Version compared to between 60 to 68% for the Home Version.

**Summary and recommendations.**

The ADHD-IV has several characteristics that make it useful for practitioners. First, the brevity of the scale allows for easy administration and scoring and, although there are no treatment sensitivity data available, demonstrates potential for repeated administration in the monitoring of treatment effects. Further research is necessary, however, to determine its efficacy as a measure of progress monitoring. Second, the manual for the ADHD-IV provides a wealth of information concerning the psychometric properties of the instrument, and these data are well organized and accessible. Further, all of the materials needed to administer the instruments, including the checklists themselves, are reproducible and included in the manual. Finally, there is good evidence for the validity of the scores derived from the ADHD-IV as meaningful indicators of ADHD symptoms. Thus, the ADHD-IV can appropriately be included as a component in the screening and multimethod phases of a school-based approach to comprehensive assessment of ADHD.
Behavior Assessment System for Children: Monitor for ADHD

The primary purpose of the BASC Monitor (Kamphaus & Reynolds, 1998) is to differentiate between the three subtypes of ADHD: predominately inattentive type, predominately hyperactive-impulsive type, and combined type. Teacher (Teacher Monitor Ratings; TMR) and parent (Parent Monitor Ratings; PMR) versions are available along with an observation component, the Student Observation System (SOS), that can be used in conjunction with the TMR. Both forms of the BASC Monitor are short and easy to complete, and the items represent general statements about key behaviors characteristic of ADHD. The BASC Monitor also includes items that identify internalizing and adaptive skill deficits to aid in the design of comprehensive treatment programs.

Although the items are highly representative of the DSM-IV criteria, some items are not scored. This occurs because the normative data for the BASC Monitor are based on the teacher and parent forms of the original Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992), which did not include certain ADHD symptom criteria of the DSM-IV. Because no new normative data were collected during the development of the BASC Monitor, the new items are not scored. Rather, these new items are examined individually when evaluating results from the BASC Monitor.

Standardization sample and norms. The manual provides norms for the ages of 4–5, 6–7, 8–11, 12–14, and 15–18 years. The authors used the same standardization data for the BASC Monitor as was used for the original BASC, and this sample was collected from 1988 to 1991. The standardization sample for the TMR consisted of 2,401 cases. A sample of 3,483 cases was collected for the PMR. The samples were weighted in order to match the sample distribution with the general population according to gender, ethnicity, and mother’s education. The standardization sample, however, is not geographically representative of the U.S. population. The North Central Region is overrepresented, whereas the Northeast Region is underrepresented in the sample. (The South Region also is overrepresented for the sample of children between the ages of 4 and 5 years.) Despite this disparity, the sample is large and representative across gender, ethnicity, and mother’s education.

Although demographic data are presented across three different age groups (4–5, 6–11, 12–18), it may have been more useful to present demographic data across the five norm groups (4–5, 6–7, 8–11, 12–14, 15–18). This approach would allow the user to assess the quality of the standardization sample for the particular age range of interest. Perhaps more significantly, creating norms based on data from the standardization sample of the original BASC presents some possible limitations to validity evidence, which are discussed in the following section.

Psychometric properties. The authors reported that three forms of reliability were examined for the BASC Monitor, including internal consistency, test-retest stability, and interrater reliability. The internal consistency for the TMR was adequate (.77–.93) across all the different subscales of the BASC Monitor. There was only one age group and subscale where the coefficient alpha dropped below .80 (internalizing problems with 4–5-year-old children). The internal consistency for the PMR was not as strong. Specifically, coefficient alphas for the hyperactivity and internalizing subscales were somewhat variable (.57 to .83) for all age groups. Thus, practitioners using the PMR should be cautious when making diagnostic decisions, especially when attempting to differentiate those with ADHD predominantly hyperactive/impulsive subtype and individuals with internalizing problems.

Test-retest data were collected from teachers and parents over a relatively broad time period (2 to 8 weeks). Stability coefficients indicated that the test-retest reliability was variable (.60 to .93) across age groups for both the TMR and PMR. Scores were less stable for ages 4–5 (.72 to .91) on the TMR, and ages 6–11 (.70 to .85) and 12–18 (.60 to .75) on the PMR. Thus, an examiner should use caution when monitoring behavior change over time with these age ranges and subscales on the TMR and PMR. An additional test-retest sample was col-
lected for the TMR over a 7-month period. The results of this analysis revealed moderate long-term test-retest stability (.54 to .82).

It is important to note that the test-retest sample included few (6.5% of TMR ages 6 to 11 sample, 0% of all other samples) individuals from Hispanic backgrounds. This is a smaller percentage than the standardization sample, which included a representative proportion of the Hispanic population (11%). In addition there were indications that interrater reliability would be reported in the manual, but no relevant tables or text appear in the BASC Monitor manual.

Kamphaus and Reynolds (1998) provided two types of evidence regarding the validity of the BASC Monitor. The first type of evidence relates to the ability of the BASC Monitor to differentiate between groups (e.g., students with different ADHD subtypes, students with and without ADHD). Several of the cited studies were not conducted using the actual BASC Monitor; rather they featured secondary analysis of data collected with the original BASC. Studies by Vaughn, Riccio, Hynd, and Hall (1997) and Ostrander, Weinfurt, Yamold, and August (1998) demonstrated at least a 1.5 standard deviation difference between ADHD and non-ADHD groups on all but one of the attention problems and hyperactivity scale comparisons. Vaughn et al. (1997) also found significant differences between children with ADHD predominantly inattentive type, ADHD combined type, and children without ADHD on the Attention Problem and Hyperactivity scales of the BASC Teacher and Parent Forms.

Correlations between the scales on the PMR/TMR and other behavior rating scales represent the second type of validity evidence. Rating scales used in the correlational analyses included the original BASC Behavior Rating Scale, Teacher Report Form and Child Behavior Checklist (Achenbach, 1991a), Conners’ Teacher and Parent Rating Scales (Conners, 1989a, 1989b), Revised Behavior Problem Checklist (Quay & Peterson, 1983), and the Personality Inventory for Children—Revised (PIC-R; Lachar, 1982). The pattern of correlations between the PMR/TMR and each of these instruments was consistent with the authors’ theoretical expectations for both the TMR and PMR, showing the highest positive correlations between attention problems and hyperactivity scales (.44–.74). Correlations between the BASC Monitor PMR and TMR and similar subscales on other instruments were high (.36–.99), particularly between the Attention Problems scale of the PMR/TMR and the Attention Problems scale of the TRF, CBCL, and original BASC. Low correlations (.05–.52) were observed between the PMR for ages 4 to 5 and the PIC-R for ages 3 to 5.

As mentioned previously, several of the studies cited to support the validity of the BASC Monitor were secondary analyses of data from the original BASC. Although these data include many of the items on the BASC Monitor, the ratings were given within the context of a longer rating scale examining multiple behavior problems. This may have affected the manner in which respondents rated certain items. For example, if a child exhibits significant somatization problems (which are included on the original BASC), the individual completing the scale may use these problems as a reference, considering the attention and hyperactivity problems as relatively less important. As a result of this limitation, studies using the final form of the BASC Monitor should be conducted. In addition, several age ranges were not included in the validity studies reported in the manual. All of the reported studies examining score profiles and differential diagnosis were limited to children age 6 to 11 years. The scale, however, is intended for use with children ages 4 to 18 years.

**Summary and recommendations.**
The purpose of the BASC Monitor includes differentiating ADHD subtypes, aiding in the design of comprehensive treatment programs and evaluating treatment effects. Test-retest stability of the BASC monitor is relatively strong, but the noted limitations (e.g., variability of the PMR test-retest stability for ages 12 to 18) require practitioners to make decisions about using the BASC Monitor for this latter purpose on a case-by-case basis. Interpretive guidelines to aid in tracking behavior change are provided, but sensitivity to behavior change by the BASC Monitor remains to be further explored.
Validity evidence presented in the manual supports the use of the TMR and PMR for children age 6 to 11; however, validity evidence was not presented regarding the use of the TMR and PMR with younger and older students. Overall, the BASC Monitor has several qualities that make it practitioner-friendly and potentially useful to monitor students’ responses to interventions, but evidence to support its use is primarily limited to children between the ages of 6 and 11.

Conners’ Rating Scale–Revised

The Conners’ Rating Scale–Revised (CRS-R; Conners, 1997) was designed to evaluate problem behaviors by obtaining reports from teachers, parents, and adolescents. Long and short versions of this norm-referenced scale are available reflecting a variety of DSM-IV criteria for childhood disorders. The primary purpose of the CRS-R is to assist in the assessment of ADHD and related problem behaviors in children and adolescents (ages 3 to 17). Directions for administration and scoring of all forms are clear and simple. All test materials are generally easy to use, appear consistent with the stated purpose of the test, and are available in a Spanish version.

Standardization sample and norms. The CRS-R was normed on several large samples of children and adolescents. The author accumulated approximately 11,000 cases during standardization and used over 8,000 cases in the normative sample. Site coordinators collected data from over 200 schools across 45 states and 10 provinces throughout the United States and Canada. The author stated that the normative sample was representative of the North American population; however, ethnic minority students were underrepresented. Specifically, parent informants who indicated the race of the target student to be African American/Black comprised only 4.3 to 4.8% of the normative sample (U.S. Census estimates for 1997 indicate that African Americans comprised 12.7% of the population). Based on 1997 U.S. Census estimates, the Hispanic population was substantially underrepresented, and the Asian population was slightly underrepresented in the normative sample. Children from special education classes were excluded from the sample. Normative data for the CRS-R are grouped by age (3-year intervals) and sex for the parent and teacher forms.

Psychometric properties. Three types of reliability information are provided for the CRS-R: internal consistency reliability, standard error of measurement, and test-retest reliability. Coefficients for all forms of the test ranged from .73 to .91, with most exceeding a moderate reliability coefficient (>.80). Coefficients for subtests specific to ADHD symptomology were from .87 to .94 for the Conners’ Parent Rating Scale Revised Long Version (CPRS-R:L); from .90 to .96 for the Conners’ Teacher Rating Scale Revised Long Version (CTRS-R:L); and from .83 to .88 for the Conners-Wells’ Adolescent Self-Report Scale Long Form (CASS:L). Alpha coefficients for the Conners-Wells’ Adolescent Self-Report Scale (CASS) were only available for a small sample of 75 adolescents. As a result of the high internal consistencies, standard errors of measurement are low, indicating minimal error in scores on the CRS-R.

Test-retest reliability was examined using an interval of 6 to 8 weeks, and stability was moderate to high across all forms. Coefficients for the long form of the parent and teacher versions ranged from .47 to .88. The range of coefficients for subtests specific to ADHD symptomology for the long form of the parent and teacher versions was the same (.47 to .88). A slightly more narrow range was evident for the short form of the parent and teacher versions for ADHD-specific subtests (.72 to .92). Coefficients for the CASS (long form) ranged from .73 to .89 for all subtests and .73 to .80 for ADHD-specific subtests.

Internal validity of the CRS-R was measured by examining the intercorrelations between the subscales to determine if they met theoretical expectations and to test the replicability of the subscale structure. Low to moderate correlations among the seven subscales of the long form across males (median $r = .34$) and females (median $r = .32$) provide evidence that the CRS-R assesses distinct dimensions of problem behavior and psychopathology. The pattern of intercorrelations was compared across gender and found to be nearly
identical. Confirmatory factor analysis supported a three-factor model for the short versions of the CRS-R, and results of additional confirmatory factor analyses were similar for the various forms of the scale (i.e., long, short, parent, teacher, and self-report).

Evidence for convergent and discriminant validity was explored through correlations among the teacher, parent, and self-report forms; correlations among the CRS-R subscales and other self-report scales; and correlations between the CRS-R and performance measures. Correlations between parent and teacher ratings generally were low to moderate. For the long form, parent and teacher correlations ranged from .12 to .47 for males and from .21 to .55 for females. Correlations between parent and teacher ratings on the DSM-IV Symptoms subscales and the Conners’ Global Index were variable (ranging from .16 to .50 for males and females). Moderate correlations were reported between the ADHD Index on the parent and teacher forms for both males and females. The pattern of correlations between parent and teacher ratings for the short form was very similar to that reported for the long form.

Low or nonsignificant correlations were reported between parent and adolescent ratings and teacher and adolescent ratings for both the long and short forms. Exceptions include moderate correlations identified between self-report and parent ratings of Cognitive Problems (.53 for males and .45 for females). Self-report and teacher ratings of Cognitive Problems also were moderately correlated (.41 for males and .40 for females). Correlations on ratings for the ADHD Index across groups ranged from .16 to .49. These findings are consistent with cross-informant results of other behavior rating scales (e.g., Achenbach, McConaughy, & Howell, 1987).

Subscales of the CRS-R appeared to correlate well with other commonly used rating scales and performance measures. The correlation between the CRS-R and the Children’s Depression Inventory (CDI) was high (.74) for cognitive problems and inattention. A significant correlation (.33) also was identified between the Continuous Performance Test (CPT; Conners, 1995) and the CPRS-R DSM-IV Symptoms Inattention subscale. A comparison of the CRS-R and the CRS-93 (an earlier version of the CRS-R) revealed moderate to high correlations between subscales for which constructs were theoretically similar (.32 to .94 for parent and teacher long forms).

Predictive validity of the CRS-R was examined by comparing a nonclinical group with a group of children diagnosed with ADHD. Prevalence data were consistent with the prevalence of the disorder in the general population (3.85% from teacher ratings and 2.30% from parent ratings). Significant differences were reported on all subscales (except Perfectionism) between the clinical and nonclinical groups for the teacher and parent ratings. A second study was conducted comparing a sample of students with ADHD and a group of children experiencing “emotional problems.” Consistent with the principal features of ADHD, the ADHD group scored significantly higher than the emotional problems group on several subscales from the parent form (Inattentiveness, Hyperactivity, Impulsivity, and other ADHD characteristics). The emotional problems group scored significantly higher than the ADHD group on the Oppositional, Perfectionism, and Social Problems subscales.

Summary and recommendations. The CRS-R is a useful tool for evaluating problem behaviors of children between the ages of 3 and 17. The teacher, parent, and/or self-report forms can be used to assist in the process of a multimethod assessment when a child is exhibiting behavioral problems including ADHD symptoms. Overall, the items are well developed and consistent with the stated purpose of the test. Some limitations were evident in the underrepresentation of certain minority groups (e.g., African American and Hispanic) in the normative samples. The addition of the self-report measure is positive; however, caution should be exercised in the use of the self-report measure given the limited evidence provided regarding its psychometric properties. The short form for all versions provides a time-efficient instrument for gathering information about a child’s ADHD-related behaviors and will be useful in treatment evaluation.
ADHD Symptom Checklist-4

The ADHD Symptom Checklist-4 (SC-4; Gadow & Sprafkin, 1997) is a 50-item norm-referenced scale designed to assess several categories of disruptive behavior, reflective of DSM-IV criteria, as well as side effects that children may experience receiving stimulant medication. The SC-4 is primarily a screening instrument for ADHD and ODD. Due to the brevity of the scale, and the inclusion of the Peer Conflict and Stimulant Side Effects Checklist, the SC-4 appears well suited for monitoring the efficacy of interventions for disruptive behavior.

Standardization sample and norms. The normative data for the SC-4 were expanded recently (Gadow & Sprafkin, 1999). A total of 1,844 parents and 2,715 teachers completed the SC-4 for children and adolescents between 3 and 18 years of age. It is important to note that much of the normative data for the SC-4 categories of ADHD and ODD were collected using other checklists developed by the same authors (e.g., Child Symptom Inventories, Early Childhood Inventories). However, a recent study by this group has suggested that responses to items on these scales are virtually equivalent (Sprafkin, Gadow, & Nolan, 2001). Normative data were collected across several geographic regions and are reported by age (3–5, 6–12, and 13–18) and sex. Minorities were somewhat underrepresented in some age groups.

Psychometric properties. The stability of the SC-4 scales is generally good. Test-retest reliability coefficients (6-week interval) for the ADHD, ODD, and Peer Conflict categories ranged between .70 and .89 for the teacher-completed checklist (Gadow, 1986a, 1986b), and between .75 and .82 for the parent-completed checklist (Pierre, Gadow, & Sprafkin, 1997). The internal consistency of the parent- and teacher-completed disruptive behavior categories is high, ranging from .92 to .95. One significant limitation is that there are no factor analytic data available to verify the internal structure of the complete SC-4. However, in a recent study, Hartman et al. (2001) investigated the factor structure of DSM-IV referenced scales including the Child Symptom Inventory, which includes items comparable to those of the SC-4 (Sprafkin et al., 2001). The scales of the SC-4 correlate well with other rating scales commonly used in the assessment of ADHD. Specifically, moderate to large positive correlations (.44–.81) have been found between the parent-completed SC-4 and relevant subscales of the Child Behavior Checklist (Achenbach, 1991a) and the Mothers’ Method for Subgrouping (Loney, 1984). Similarly, moderate to large positive correlations (.45–.88) were found between relevant subscales of the Teacher Report Form (Achenbach, 1991b) and the IOWA Conners’ Teacher’s Rating Scale (Loney & Milich, 1982) and the teacher-completed SC-4 (Gadow & Sprafkin, 1997).

Predictive validity data for teacher- and parent-completed checklists are available for preschool- and elementary school-aged samples. Data concerning the predictive validity for ratings of adolescent behavior are available for the parent completed checklist only. Scores on the parent- and teacher-completed ADHD-Inattentive Type, ADHD-Combined Type, ODD, and Peer Conflict categories were significantly different between outpatient clinic and normal school samples across age groups. Differences between groups on the ADHD-Hyperactive-Impulsive Type, although significant for males across age groups, were not significant for teacher reports of elementary school-aged females or parent reports of adolescent females (Gadow & Sprafkin, 1997; Sprafkin & Gadow, 1996).

The diagnostic utility of the SC-4 was investigated for three age groups (preschool, elementary, and adolescent) by comparing clinic diagnoses to screening cutoff scores on the SC-4 ADHD and ODD scales. The indices of sensitivity (the degree to which a scale correctly identifies children with a given disorder) and specificity (the degree to which a scale correctly identifies children without a given disorder) were used to assess the diagnostic utility of the SC-4. Agreement between the SC-4 and clinic diagnosis appears to vary by age group, category, and informant. In general,
sensitivity for the ADHD scales ranged from .81 to .85 (specificity from .60 to .85) for parent reports and .61 to .89 (specificity from .57 to .94) for teacher reports.

Several studies have investigated the treatment sensitivity of the SC-4 categories (e.g., Gadow, Nolan, Sved, Sprafkin, & Paolicelli, 1990; Gadow, Sved, Sprafkin, Nolan, & Ezor, 1995; Sved, Gadow, & Paolicelli, 1989). For the parent- and teacher-completed ADHD, ODD, and Peer Conflict scales, statistically significant differences have been found between placebo and three doses of methylphenidate, albeit not necessarily between individual doses. Scores on these scales were generally 2–3 times higher during placebo than those gathered during the highest dose of methylphenidate (.5 mg/kg). Although several significant differences have been found between doses on both the parent- and teacher-completed Stimulant Side Effects Checklist, the trend in scores for teachers reported by Gadow and Sprafkin (1997) indicated a decrease in side effects from placebo to higher doses. There is no clear explanation of this finding at this time; however, it may be an artifact of the medication acting to reduce informants’ awareness of a child’s symptoms.

**Summary and recommendations.**

The SC-4 offers a broad assessment of disruptive behavior including ADHD symptoms and may be especially useful for monitoring the effects of stimulant medication in school settings. The lack of factor analytic evidence supporting the internal structure of the SC-4 item groupings is a significant limitation. Both the DSM-IV categories and the Peer Conflict and Stimulant Side Effect categories are clinically, rather than empirically, derived. Nevertheless, the SC-4 appears to be a relatively easy instrument to use and score, with validity evidence to support its use for screening and/or monitoring purposes.

**The Attention Deficit Disorders Evaluation Scale—Second Edition**

The Attention Deficit Disorders Evaluation Scale (ADDES [2nd ed.]; McCarney, 1995) is a norm-referenced behavioral rating scale, composed of separate forms for both teachers and parents. Normative data are available for students aged 4 to 19 (School Version) and ages 3 to 19 (Home Version). The author identifies several uses of the ADDES including: screening for ADHD; assisting in the diagnosis of ADHD; and aiding in the development of individualized goals, objectives, and intervention strategies for the home and school environments.

Particular strengths include straightforward directions regarding completion of the protocol, clear items qualified with examples, and a profile sheet including a visual chart upon which standard scores and percentile ranks can be plotted. The technical manuals accompanying the ADDES home and school versions are also clear, comprehensive, and well written. The publisher of the scale also offers a Secondary school version; forms for comparing ADDES results to DSM-IV criteria; intervention manuals for elementary age, secondary age, and parents; rating forms in Spanish; and computerized scoring forms.

**Standardization sample and norms.**

In 1994, the ADDES School Version was standardized on a total of 5,695 students (ages 4–18) rated by their teachers, and the Home Version was standardized on 2,415 children (ages 3–20) rated by their parents. Both versions include normative data for children with and without ADHD. According to the author, demographic characteristics of the standardization sample represent national percentages of sex, residence, race, geographic area, and occupation of parents based upon the 1992 Statistical Abstract of the United States. Scores were found to vary significantly by sex (i.e., males exhibited more inappropriate behaviors than females, resulting in higher scores), and on some subscales by age (i.e., scores generally decreased with age). Thus, the author developed 10 standardization groups providing separate norms for males and females across five age ranges. Differences between racial groups were not investigated.

**Psychometric properties.** The author presents detailed information about the psychometric properties of the Home and School
Versions. Reliability evidence includes internal consistency, test-retest, and interrater reliability. The Kuder-Richardson 20 (KR20) formula was used to assess internal consistency. The KR20 values of the Inattentive and Hyperactive/Impulsive subscales all exceeded .90 (School) and .95 (Home). Test-retest stability for both versions was excellent (.88–.97) and was assessed using randomly selected groups of children from within the normative population (n = 148 for Home Version; n = 481 for School Version) 30 days after their initial rating. Interrater reliability for the school Version included 462 students, each rated by a pair of educators with similar knowledge of the students. Interrater reliability for the Home Version was conducted with 86 children rated by both parents. In both cases, strong interrater reliability was obtained (.85 and .82 for the school and home versions, respectively).

Validity data also were presented for the two versions of the ADDES. Content validation of items occurred through expert (i.e., diagnosticians, educators, parents) review, aligning items with APA and DSM-IV criteria, an initial field test, and a second review of retained items by expert judges. Confirmatory factor analyses supported a two-factor model (Inattentive and Hyperactive/Impulsive) for both versions. Finally, diagnostic validity was supported through the comparison of randomly selected children from the standardization sample to children who had been diagnosed with ADHD by outside evaluations. For both the Home and School Versions of the ADDES, differences in mean scores were statistically significant (p < .001) and in the expected direction across ADHD subscales. Unfortunately, these examinations of discriminant validity were performed on a sample of children and youth between the ages of 4.5 and 19 years. Hence, it is unclear whether the scale demonstrates discriminative power for each of the age ranges collapsed in these analyses (e.g., preschool, elementary, adolescent).

As evidence of construct validity, the author utilized several samples of children previously identified with ADHD to compare the ADDES home and school versions to other validated measures with similar constructs. These measures included the Conners’ Teacher Rating Scale-28 and 39 (CTRS-28, CTRS-39; Conners, 1989a), Conners’ Parent Rating Scale-48 and 93 (CPRS-48 and CTRS-93; Conners, 1989b), ACTeRS (Ullman, Sleator, & Sprague, 1991), Children’s Attention and Adjustment Survey (CAAS; Lambert, Hartsough, & Sandoval, 1990), and the Child Behavior Checklist and Teacher Report Form (CBCL & TRF; Achenbach, 1991a, 1991b). Demographic information about the samples is limited to children’s ages, which ranged from 4 to 19 years. In each study, moderate to strong correlations (range = .42–.90) were found between relevant subscales of the ADDES versions and the comparison measures listed. However, caution should be exercised when evaluating the results of these respective studies, as the sample sizes were relatively small (n = 42–79).

Summary and recommendations.

The ADDES offers a relatively quick and easy-to-use instrument for screening for ADHD. The protocols and manuals accompanying the Home and School Versions are clear, comprehensive, straightforward, and much improved from the original version. The ADDES possesses adequate to excellent overall reliability evidence and has a solid theoretical foundation. The author provides evidence for the validity of the scores, demonstrating that the ADDES correlates well with other measures of ADHD and related behavioral disorders/deficits. A limitation of the validity evidence is the relatively small sample sizes utilized to conduct the analyses for the validity studies. An additional limitation of the validity evidence is that differences in scores across various racial groups were not examined in the standardization sample. In sum, the ADDES is a user-friendly scale that has evidence to support its use as an initial component of a comprehensive assessment for ADHD.

ADD-H Comprehensive Teacher’s Rating Scale

The ADD-H Comprehensive Teacher’s Rating Scale (ACTeRS; Ullmann, Sleator, & Sprague, 1997) behavioral rating scales were
developed to help identify children with attention disorders and monitor the efficacy of treatment. The three forms of the current instrument are a revision of an earlier version published in 1991. The current review examines the teacher and parent forms of the ACTeRS, which were developed together. The ACTeRS teacher and parent forms are short and easy to complete. The items are written as general statements about behaviors that may change with intervention, making the instrument potentially useful for evaluating treatment effectiveness over time. In addition, the brevity of the scale makes it convenient for frequent administration.

**Standardization sample and norms.** Initial norms for the ACTeRS were based on a sample consisting of 1,339 kindergartners through fifth graders. Normative data for the current version of the ACTeRS were collected in 1989, and included a sample of 2,362 students from 23 schools, rated by 84 teachers. Parent ratings were not collected in the most recent standardization efforts. Instead, test-equating techniques were employed to ensure that the percentile rank scores on the parent form were similar to the distribution of scores on the teacher form.

The most recent standardization sample of the ACTeRS expanded the previous norms to include students in kindergarten through eighth grade. A majority of the students included in the most recent sample, however, were from the middle grades. Specifically, of the 2,362 students included in the normative sample, those in Grades 6 (n = 518), 7 (n = 448), and 8 (n = 493) comprised 62% of the total sample. Approximately 150 students from each of the lower grades were included in the most recent normative sample. This approach allowed the authors to create current norms, while examining the stability of the previous norms.

The normative sample for the ACTeRS is large; however, it is impossible to determine the appropriateness of the sample due to the lack of demographic information reported in the manual. In addition, it would have been more accurate to develop norms for the parent form using ratings completed by parents rather than a test-equating procedure.

**Psychometric properties.** Internal consistency, test-retest stability, and interrater reliability were reported in the ACTeRS’ manual. Internal consistency estimates for both forms of the ACTeRS were reported for both the previous and current samples indicating high coefficients ranging from .92 to .97. The test-retest correlations (n = 80) for the teacher form demonstrated acceptable stability, ranging from .78 to .82. Interrater reliability was based on ratings completed by only two teachers for 124 boys and girls. Although estimates were acceptable, ranging from .51 (Social Skills subscale) to .73 (Hyperactivity), the small teacher sample may have inflated the coefficients. Overall, the data provide mixed evidence regarding the reliability of scores from the ACTeRS. In addition, given that one of the primary goals of the ACTeRS is to assess treatment effectiveness, it is particularly important for the authors to examine test-retest stability data. The data, however, are not reported for the parent form of the ACTeRS.

The authors indicate that two primary sources of validity evidence are included in the ACTeRS manual: correlations between parent and teacher ratings and a discriminant function analysis. However, no correlations between parent and teacher ratings appear in the manual. Fortunately, data on the ability of the ACTeRS parent form to discriminate between students identified with ADHD and students without ADHD were provided in the manual. The sample for this analysis included a total of 477 children, 256 of whom had been diagnosed with an attention disorder. Significant mean differences were reported (p < .01) between diagnosed and nondiagnosed children regarding attention and hyperactivity. Additional validity studies cited in the manual support the utility of the ACTeRS in discriminating students identified with ADHD from those identified with learning disabilities (Peoples, 1989) and monitoring treatment effects (Brigman, 1991; Metzger, 1992; Ullman & Skeator, 1985, 1986). The majority of these latter studies examined the ability of the ACTeRS to identify changes in behavior based on the implementation of medical, curriculum, and
cognitive-behavioral interventions used to reduce symptoms of ADHD.

Summary and recommendations. Despite some positive qualities (i.e., brevity, ease of use, and potential to monitor treatment efficacy), the limitations of the ACTeRS make it difficult to recommend its use in practice as an instrument for screening or identification of ADHD. Some of these limitations include a lack of evidence to support stated goals, vague and limited information regarding the standardization sample and normative data, extrapolation of scores for the parent form, and insufficient research with the current versions of the scale. A notable theoretical limitation appears to be the lack of connection between the items on the ACTeRS scales and the DSM-IV diagnostic criteria. Although some validity evidence is reported in the ACTeRS manual, the evidence is not sufficient to justify its intended uses.

Discussion

Conclusions and Overall Recommendations for Practitioners

The primary purpose of this review was to evaluate critically six published rating scales designed to assess behaviors associated with ADHD. Table 3 provides a summary of the strengths, limitations, and recommended uses of each scale as it fits within a comprehensive school-based approach to assessment of ADHD. Recommendations for practice based on this review are organized according to three broad categories of the assessment process: screening, multi-method or comprehensive assessment, and progress monitoring or response to treatment.

Screening. With the exception of the ACTeRS, all of the reviewed rating scales could be used for screening. The CRS-R (short form) and ADHD-IV are ideal choices given their brevity, focus on the primary symptoms associated with ADHD, and strength of psychometric evidence specific to the purpose of screening. There is strong evidence to support the use of the SC-4 in the screening process. Sufficient evidence exists to support the use of the ADDES and BASC-Monitor for screening purposes, but their primary strengths are related to other assessment goals. The serious limitations noted regarding the ACTeRS prevent recommending its use for screening (or other purposes) at this time.

Multi-method assessment/comprehensive evaluation. The ADHD-IV and the CRS-R are ideal choices for use in the multimethod assessment phase. Both measures are well validated and contain items that closely approximate the DSM-IV symptom criteria for ADHD (CRS-R long form only). Both have Spanish translations, but their appropriateness for use with culturally diverse students remains an empirical question. Both measures can also facilitate the treatment development process by highlighting certain items to consider for targeting change. The comprehensive nature of the long version of the CRS-R is reflective of DSM-IV symptom criteria for childhood disorders and provides data regarding areas of potential comorbidity requiring additional assessment. The ADDES may also be considered for this type of assessment given its comprehensiveness and precise wording of items, which may facilitate the selection of specific target behaviors.

Progress monitoring. With the exception of the SC-4, all of the rating scales in this review had limited evidence regarding sensitivity to changes in behavior (e.g., resulting from treatment). In addition to providing such evidence, the SC-4 includes a side effects scale that makes it the best choice for monitoring the impact of pharmacological interventions. The brevity of the ADHD-IV and CRS-R (short form) make them convenient for treatment monitoring; however, additional evidence is necessary to justify their use for this purpose. Additional evidence is necessary to support the use of the BASC Monitor for this purpose as well. Finally, the limited empirical evidence of the ADDES, in addition to its scaling (e.g., a rating of 4 indicates that the behavior occurs one to several times per hour) severely limits its use for treatment monitoring.

Future Research Directions

Several directions for future research are evident. First, given the lack of empirical sup-
Implications for Practice

It is important for school-based practitioners to note that, regardless of the purpose of assessment, all rating scales included in this review lacked conclusive evidence regarding the appropriateness of their use with students from culturally diverse backgrounds. For this reason, utilizing multiple methods of assessment cannot be overemphasized. A best practices approach for assessment of culturally diverse students requires adherence to a problem-solving framework and obtaining culturally relevant information from multiple sources through a variety of methods (Ridley, Li, & Hill, 1998). Given the limited normative data for these populations among the reviewed instruments, school psychologists must include additional direct methods of assessment to determine severity and pervasiveness of symptomology as well as response to treatment. Indeed, the upcoming reauthorization of the Individuals with Disabilities Education Act (IDEA) may result in a shift towards methodologies focusing on early identification and prevention efforts as well as the importance of establishing responsiveness to treatment (Pasternack, 2002).

Although all of the reviewed rating scales offer a brief and convenient method for obtaining information about a child’s ADHD symptoms, only five had sufficient evidence to justify their use for inclusion in a school-based assessment paradigm. The limitations associated with behavior rating scales preclude their use in isolation from other methods of assessment such as interviews with parents and teachers, direct observations, review of school records, and academic skills assessment (DuPaul & Stoner, 1994). For example, misperceptions and misunderstandings of items have been associated with the use of rating scales (Merrell, 2000). However, rating scales are efficient, allow for a comparison of students with same-aged peers, and provide a means for aggregating information across situations and time. They do not provide exhaustive information about the child, environmental variables, or information relevant to response function. Thus, ratings from behavior scales have limited capacity to render formal diagnosis and are inappropriate for use in conducting functional behavioral assessments specifically related to hypothesis testing.

Personal preference for a particular scale should never determine its selection for assessment (Landau & Burcham, 1995). It is hoped that the information in this review will assist school-based practitioners with selecting the most appropriate tool based on characteristics of the child to be assessed, the primary purpose of the assessment, and the adequacy of the psychometric property of the assessment device.

References


nal of Psychopathology and Behavioral Assessment, 20, 57-81.


Lisa Marie Angello is a doctoral candidate in the School Psychology program at Lehigh University. She is currently completing a predoctoral internship at the University of Florida, Family Safety Behavior Analysis Initiative in Gainesville, FL. Her research and professional interests include assessment methodologies possessing treatment validity such as curriculum-based measurement and functional behavioral assessment, and delivery of psychological services through a behavioral model of consultation.

Robert J. Volpe is a doctoral candidate in the School Psychology Program at Lehigh University. He is currently completing a predoctoral internship at Centennial School in Bethlehem, PA. His research and professional interests include academic interventions for students with attention-deficit/hyperactivity disorder and related behavior disorders and early intervention for the prevention of behavior and academic problems.

James Clyde DiPerna, Ph.D., is an Associate Research Scientist in the Wisconsin Center for Education Research at the University of Wisconsin-Madison. His research interests include academic and emotional competence, school-based prevention programs, and the use of technology to facilitate learning.
Sammi Gureasko-Moore is an advanced doctoral student in the School Psychology Program at Lehigh University. She is currently completing her doctoral internship at the Virginia Beach City Public Schools, in the public health track. She has extensive experience working with children and adolescents with ADHD. Currently, Mrs. Gureasko-Moore is preparing her dissertation, which will further investigate the effects of self-management interventions for adolescents with ADHD.

David Gureasko-Moore is an advanced doctoral student in the School Psychology Program at Lehigh University. He is currently completing his doctoral internship at the Virginia Beach City Public Schools, in the gifted education track. His research interests include creating partnership between health care institutions and schools. Mr. Gureasko-Moore is completing his dissertation, which examines the role of the school psychologist in medication monitoring for students with ADHD.

Michelle Raffa Nebreg is an advanced doctoral student in the School Psychology Program at Lehigh University and is currently employed as a school psychologist in the Charlotte-Mecklenburg County, NC Schools. Her research and professional interests include school-based medication monitoring for ADHD, and active training models for teachers to prevent academic and behavior problems.

Kenji Ota is an advanced doctoral student in the School Psychology Program at Lehigh University and is currently employed as an athletic director for a private school in the Philadelphia suburbs. Mr. Ota is completing his dissertation investigating a math intervention for children with ADHD.