(false negatives). Again, we must caution against the use of this test for the diagnosis of ADHD.

Rey–Osterrieth Complex Figure Drawing

The Rey–Osterrieth Complex Figure Drawing (see Lezak, 1995) is a paper-and-pencil task requiring planning and visual–spatial–constructional abilities, and is sensitive to deficits from frontal lobe injuries. The task requires the subject to copy a complex geometric shape. The Webber and Holmes (1985) scoring procedure is often used, yielding scores for organization age (five levels) and style (four categories). Several studies of children with ADHD have shown that they perform this test more poorly on average than do non-disabled children (see Chapter 3, this volume). However, effect size differences from control groups are mere 0.24–0.26 (Frazier et al., 2004). And once again, Barkley and Grodzinsky (1994) found that although non-average test scores accurately predicted the presence of ADHD 100% of the time, average test scores accurately predicted the absence of disorder only 30% of the time, and the false-negative rate was a stunning 96%. Overall, the test accurately classified only 52% of the children. Therefore, we urge clinicians not to employ this test for diagnostic purposes concerning ADHD.

Trail Making Test (Parts A and B)

The Intermediate version of the Trail Making Test from the H-R is frequently used with children. It comprises two parts: A and B (Reitan & Wolfson, 1985). In Part A, the subject connects a series of numbered circles distributed arbitrarily on a page. Part B comprises circles that contain letters or numbers scattered randomly across the page; the subject is to alternate connecting numbers and letters in ascending order until the end of the sequences. The scores are the time taken to complete each part by the subject. Some studies have found this test to be useful for differentiating groups of children with ADHD from control groups (Barkley et al., 1992). The average effect size for Part A in 13 studies was 0.40, while that for Part B in 14 studies was slightly higher, being 0.59 (Frazier et al., 2004). Barkley and Grodzinsky (1994) found that the test as a whole accurately predicted presence of disorder 48–71% of the time, accurately predicted absence of disorder just 51% of the time, and had false-negative rates of 80–82%. Overall classification accuracy was just 54%. Here, then, is another test that we must recommend against for diagnosing ADHD in children.

Continuous-Performance Tests

The most popular and widely studied form of testing for use in ADHD evaluations is based on a paradigm called the CPT (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956). Although the CPT has been administered with many variations (e.g., visual, auditory, numbers, and characters), the most common one requires the youngster to observe a screen while individual letters or numbers are projected onto it at a rapid pace (typically at one per second). The child is told to respond (e.g., to press a button) when a certain stimulus or pair of stimuli in sequence appears. The scores derived from the CPT are the number of correct responses, number of target stimuli missed (omission errors), and number of responses following nontarget or incorrect stimuli (commission errors). The latter score is presumed to tap both sustained attention and impulse control, whereas the two former measures are believed to assess sustained attention only (Sostek, Buchsbaum, & Rapoport, 1980).

Researchers have been examining versions of the CPT paradigm for almost 40 years. A wide-ranging literature has shown it to be the most reliable of psychological tests for discriminating groups of children with ADHD from non-disabled children (Corkum & Siegel, 1993). Average effect size across 40 studies using the commission error score was 0.55, while that for omissions was 0.66 (33 studies), and that for correct hits was 1.00 (19 studies) (Frazier et al., 2004); these were nearly twice the effect sizes found for other tests, such as the Rey–Osterrieth or the WCST (see above). CPTs are also sensitive to stimulant drug effects among children and adolescents with ADHD (Coons, Klorman, & Borgstedt, 1987; Fischer, 1996; Garfinkel et al., 1986). Although concerns abound about its actual discriminative ability and ecological validity (Barkley, 1991), the CPT nonetheless is the only psychological measure that seems to directly assess the core symptoms of the disorder—namely, impulsivity and inattention. Moreover, the CPT assesses these dimensions without undue contamination from other cognitive factors, such as conceptual abil-
ity, visual scanning, and so on. In all its embodiments, the CPT places relatively little demand on subjects other than to sustain attention and to refrain from responding except in special circumstances. Finally, use of the CPT may be especially important when clinicians are assessing adults or those suspected of malingering. Quinn (2003) found that college undergraduates asked to simulate malingering were able to do so easily on rating scales, but not on a CPT.

The CPT serves as the paradigm for several commercially available performance measures, including the Conners CPT (Conners, 1995) and Conners CPT II (Conners & MHS Staff, 2000); the Gordon Diagnostic System (GDS; Gordon, 1983); the Test of Variables of Attention (Greenberg & Kindschi, 1996); and the Intermediate Visual and Auditory Continuous Performance Test (Sandford, Fine, & Goldman, 1995). Each format requires the child to respond to certain signals embedded in a series of irrelevant targets. Whereas each measure is a CPT, the tasks vary considerably along what would seem to represent important dimensions, including the length of the task, the type of stimulus, the duration between stimuli, and the instructions to the subject. For example, one measure requires the child to respond based on the position of a certain graphic for 23 minutes; another presents numbers on a screen for 9 minutes; and still another presents a combination of numbers both visually and auditorily. Typically, children sit passively observing (or listening) to the presentation of nontarget stimuli and then must respond (often manually) to the occasional target stimulus. The CPT that differs most from the traditional paradigm was developed by Conners (1995). In this test, the child is told to press a button on each trial until the target appears, at which time the child is to inhibit responding; a different form of response inhibition is thus required.

Which CPT should a clinician choose? Because no data have been published on head-to-head comparisons for reliability, validity, or clinical utility, we cannot offer an empirically based opinion. It is therefore unclear whether these measures would differ from one another in their contributions to accurate identification, ruling out alternative explanations, or confirming comorbid conditions. It is also uncertain whether diagnosis would be more accurate or productive if some combination of these tests were administered. For example, some investigators say that attention should be assessed both in visual and in auditory modalities. Although some data suggest that children generally find auditory versions more difficult than visual ones (Aylward & Gordon, 1997), no one has presented data that show a superior form of one format over the other. Paraphrastically, based on Barkley's theory of ADHD (see Chapter 7, this volume), it would be unlikely that a child would suffer problems with response inhibition in just one sensory modality.

In the absence of Consumer Reports-type comparisons, decisions must be based on the sort of parameters mentioned earlier in this chapter: practicality, robustness of standardization, reliability of administration, and the extent to which the technique has been scrutinized scientifically for its potential contributions (or lack thereof). Unfortunately, here again we are constrained by a literature that is limited in scope and depth. Although the past several years have witnessed more research examining the psychometric properties of CPTs, a review by Nichols and Waschbusch (2004) concludes that still more evidence of validity is needed before CPTs can reach their potential for high clinical utility. Limiting progress has been diversity among studies in criterion measures. Also, not all studies have assessed reliability, practicality, and standardization.

One measure that has published information available on all these relevant dimensions was developed by one of us (Gordon). The GDS has been used extensively in research and clinical practice. It is a portable, solid-state, child-proofed computerized device that administers a 9-minute vigilance task wherein the child must press a button each time a specified, randomly presented numerical sequence (e.g., a 1 followed by a 9) occurs. Another version of this task presents random distracters on either side of the center target. Normative data are available for more than 1,000 children on the mainland United States ages 3–16 years (Gordon & Mettelman, 1988) and for Puerto Rican children (Bauermeister, Berrios, Jimenez, Acevedo, & Gordon, 1990). Norms are also available for an auditory version of this task (Gordon, Lewandowski, Clonan, & Malone, 1996). Gordon's CPT has been found to have satisfactory test–retest reliability (Gordon & Mettelman, 1988), to correlate modestly but significantly with other laboratory measures.
of attention (Barkley, 1991), to discriminate groups of children with ADHD from nondisabled children (Barkley, DuPaul, & McMurray, 1991; Gordon, 1987; Grodzinsky & Diamond, 1992; Mariani & Barkley, 1997), and to be sensitive to moderate to high doses of stimulant medication (Barkley, Fischer, Newby, & Breen, 1988; Barkley, DuPaul, & McMurray, 1990; Fischer & Newby, 1991; Fischer, 1996; Rapport, Tucker, DuPaul, Merlo, Stoner, 1986). The GDS is useful in the assessment of children with hearing impairment (Mitchell & Quittner, 1996). A body of literature also suggests that poor GDS performance is tied to other neuropsychological measures (Grant, Ilai, Nussbaum, & Bigler, 1990) and to general academic underachievement (Aylward et al., 1997; Billings, 1996; Gordon, Mettellman, & Irwin, 1994).

One study (Mayes, Calhoun, & Crowell, 2001) investigated the clinical validity of the GDS, using a DSM-based assessment as the criterion measure. Large differences were found between groups with and without ADHD in GDS composite standard scores (d = .92) and IQ-GDS discrepancies (d = 1.08). More importantly, a discrepancy score cutoff formula (a 13-point discrepancy between GDS composite and IQ) resulted in a sensitivity of 90% and a specificity of 70%.

As with the interpretation of any psychological test for the diagnosis of ADHD, interpreting the data from traditional validity studies of CPTs can be vexing, because no gold standard is available to use for comparison with these tests (Gordon, 1993). Nonetheless, if we use either a high score on ADHD-related behavior rating scales or DSM clinical diagnoses as benchmarks, the GDS appears to accurately classify 83–90% of children with abnormal scores as having ADHD (Barkley & Grodzinsky, 1994). Its classification accuracy for children with average scores as not having the disorder is, however, 59–61%. Moreover, a range of 15–52% has been found for its false-negative rate (i.e., children who are rated by parents or teachers as having ADHD, but who obtain average scores on the test) (Gordon, Mettellman, & DiNiro, 1989; Barkley & Grodzinsky, 1994; Trommer, Hoeppner, Lorber, & Armstrong, 1988). Therefore, as with all the neuropsychological tests discussed here, if a child performs well on this measure, it does not indicate that the child is nondisabled or without ADHD. But it may have some diagnostic significance when a child otherwise considered to have ADHD exhibits average performance on the GDS. Research by Fischer et al. (1995) shows that these children are typically rated as less impaired, more likely to show comorbid internalizing problems, and less likely to be prescribed stimulant medications. Data such as these suggest that, if anything, GDS performance might represent an indication of severity.

The false-positive rate of the instrument is good to excellent, with 2–17% of nondisabled children being classified as having ADHD (Gordon et al., 1989; Barkley & Grodzinsky, 1994). These kinds of data may be helpful in cases for which objective confirmation of the diagnosis is important. But even here, the GDS cannot be used for objective disconfirmation of the disorder, given the rate of false negatives found in some studies. Once again, the presence of a nonaverage score probably (but not necessarily) indicates the presence of disorder, whereas the presence of an average score must go uninterpreted, as it may not indicate the absence of disorder. As with rating scales, the test provides one source of information to be integrated with other sources in reaching a final diagnostic decision.

Another popular measure, the Conners CPT (Conners, 1995), has enjoyed some scientific attention. Two such studies focused specifically on its diagnostic utility in ADHD assessments—one study with children (McGee, Clark, & Symonds, 2000) and another with adults (Epstein, Conners, Sitarenios, & Erhardt, 1998).

McGee et al. (2000) investigated associations between Conners CPT scores and several other measures, including parent and teacher ratings as well as neuropsychological and achievement tests. These researchers reported several positive findings concerning the Conners CPT, including its lack of relation to age, order and fatigue effects, or peripheral motor skill. However the Conners test's overall index failed to relate to parent and teacher ratings, and only slightly over half of those participants who met criteria for ADHD “failed” this CPT (i.e., obtained a total index score of over 11). Furthermore, the test demonstrated poor discriminant validity, in that children with a reading disability actually performed more poorly than children with ADHD.

Epstein et al. (1998) conducted a similar study, using adult participants who were given
a semistructured interview to classify them as having ADHD or not, and who then were administered the Conners CPT. Importantly, these investigators created dependent variables from this CPT using signal detection theory (e.g., d'), and also used raw scores such as reaction time, rather than using cutoff scores or total index scores. Even so, none of the Conners CPT scores correlated significantly with ADHD symptoms as measured by the semistructured interview. When scores on the three most discriminating aspects of this CPT were compared with initial classifications, the test's sensitivity was only 55%, although its specificity was somewhat better (76.4%).

A more recent study assessed the newly revised version of the Conners CPT (the Conners CPT II; Conners & MHS Staff, 2000) in a school-based sample by Weis and Totten (2004). Like McGee et al. (2000) and Epstein et al. (1998), Weis and Totten's results cast doubt on the utility of the Conners CPT II, extending previous findings to the new version. These investigators found mostly nonsignificant relationships between CPT II performance and three other kinds of measures (parent ratings, teacher ratings, and classroom observations). Furthermore, Weis and Totten questioned the discriminative validity of the CPT II, due to a significant negative correlation between IQ (as assessed by the Kaufman Brief Intelligence Test) and omission scores on the CPT II. The authors interpret this finding, in the context of other results, as suggesting that the CPT II may measure letter recognition skills or phonological awareness rather than impulsivity or inattentiveness per se.

Two recent studies by Conners and his colleagues contain interesting information about the properties of the Conners CPT II as well as newer normative data, but neither paper allows for direct inferences concerning diagnostic utility. Conners, Epstein, Angold, and Klaric (2003) present interesting analyses of this CPT, looking at age x gender interactions, and also varying the interstimulus interval. Epstein et al. (2003) used generalized estimating equation statistics to show strong associations between various indices of Conners CPT II performance and ADHD symptoms, but since the latter variable is continuous, we are not able (given only the published data) to calculate easily the sensitivity and specificity of the positive and negative predictive power of this CPT for ADHD status.

Other Motor Inhibition Tasks

Two other motor inhibition tasks show promise as aids in diagnostic assessment. In the go/no-go task, an individual responds to a certain class of stimuli (e.g., leftward-pointing arrows flashed on a screen) by making some motor response (e.g., pressing a button, squeezing a dynamometer), but withholds any response to another class of stimuli (e.g., upward- and downward-pointing arrows flashed on a screen). In a similar second task, the stop signal task (SST), an individual is asked to press one of two computer keys, depending on which of two stimuli is shown on the screen—at least when a certain tone sounds before the stimulus is presented, in which case neither key should be pressed. These two tasks are both presumed to measure executive functioning in the motoric domain (Zelazo & Müller, 2002).

Recent meta-analyses (Frazier et al., 2004; Hervey, Epstein, & Curry, 2004) have found moderate to large effect sizes (0.55–0.66) when groups with ADHD and control groups are compared on the SST paradigm. In their review, Nichols and Waschbusch (2004) singled out the SST as one of the “most promising” developments in ADHD assessment, but noted that much of the convergent validation has related SST performance to other laboratory measures, rather than to tasks in more naturalistic settings. Furthermore, Nichols and Waschbusch found inconsistency across studies comparing the performance of individuals with ADHD to those in other clinical groups; the differences tended to be small and not always in the same direction.

The go/no-go task, though it comes from a rich intellectual heritage (e.g., Luria, 1966), boasts less research in clinical populations. Most of that research has dealt either with schizophrenia or with the electrophysiological underpinnings of performance on the task. The one recent study examining the discriminative validity of the go/no-go task in ADHD assessments (Berlin, Bohlin, Nyberg, & Janols, 2004) combined performance on this measure with results from eight other neuropsychological measures. There was a statistically significant difference between participants with ADHD and controls on the go/no-go task, but in a logistic regression model used to predict group status, performance on the go/no-go task did not contribute significantly, and the predictive power of the regression model did not decline.
when the go/no-go task was dropped as a variable.

**Cancellation Tasks**

Several paper-and-pencil versions of CPTs have been used as methods of assessing attention. These tasks typically involve having a child scan a series of symbols (letters, numbers, shapes) presented in rows on sheets of paper. The child is typically required to draw a line through or under the target stimulus, using a pencil. One such task, which has shown promise in discriminating children with ADHD from nondisabled children, is the Children's Checking Task (CCT; Margolis, 1972). The CCT consists of seven pages with a series of 15 numerals per row printed in 16 rows on a page. A tape recorder reads off the numbers in each row, and the child is required to draw a line through each number as it is read. Discrepancies between the tape and printed page are to be circled by the child. There are seven discrepancies between the tape and printed pages. The CCT takes about 30 minutes to complete. Scores are derived for errors of omission (missed discrepancies) and errors of commission (numbers circled that were not discrepancies). Brown and Wynne (1982) found that the task discriminated groups of children with ADHD from those with reading disabilities. The CCT correlates modestly but significantly with other measures of attention (Keogh & Margolis, 1976), often to a larger degree than any of the other laboratory measures (see Barkley, 1991). Perhaps this is because it is somewhat similar to the academic accuracy demands made during work that children must do in the classroom. Nevertheless, its sensitivity to ADHD symptoms requires more research and replication of these initially promising results before it can be recommended for clinical practice.

**Matching Familiar Figures Test**

This Matching Familiar Figures Test (MFFT; Kagan, 1966) has a lengthy history of use in research studies investigating impulse control in both nondisturbed and disturbed children and adolescents. This match-to-sample test involves the examiner's presenting a picture of a recognizable object to the youngster, who must choose the identical matching picture from among an array of six similar variants. The test includes 12 trials, with scores derived for the mean time taken to the initial response (latency) and the total number of errors (incorrectly identified pictures). A longer version of the MFFT employing 20 stimulus trials (MFFT-20) has been developed (Cairns & Cammock, 1978); it is purport to achieve greater reliability among older children and adolescents (Messer & Brodzinsky, 1981). Unfortunately, more recent research on the original test has often failed to find significant differences between children with ADHD and nondisabled children (Barkley et al., 1991; Fischer et al., 1990; Milich & Kramer, 1984). A meta-analysis of 11 studies found an average effect size of 0.27 for the timing score, but a more satisfactory result of 0.60 for the errors score (Frazier et al., 2004). The MFFT has also shown conflicting and often negative results in detecting stimulant drug effects in children with ADHD (Barkley, 1977; Barkley et al., 1991). Furthermore, norms for the adolescent population are not currently available for either the MFFT or the MFFT-20, thus limiting their use as diagnostic measures. Evidence for acceptable positive and negative predictive power is lacking, and no new research on this measure could be located since the publication of the preceding edition of this text that would alter these conclusions. The MFFT appears to have fallen into both clinical and scientific disuse. Consequently, we do not endorse this instrument for use in clinical practice in making diagnostic decisions about ADHD in children.

**GDS Delay Task**

The GDS Delay Task, a part of the GDS discussed above (Gordon, 1983), is a measure of response inhibition. It utilizes a paradigm of direct reinforcement of low rates of behavior. The child sits before the portable, computerized GDS device and is told to wait a while before pressing a large blue button on the front panel of the device. Children are told that if they have waited long enough, they will earn a point when they push the button. If they press it too early, they must wait a while before pushing the button again. Cumulative points are scored on a digital counter on the face of the device. The child is not informed of the actual delay required to earn reinforcement (6 seconds). The test lasts 8 minutes and has normative data for more than 1,000 children. Initial evidence (Gordon, 1979; Gordon & McClure,
1983; McClure & Gordon, 1984) indicated that the test discriminated groups of children with ADHD from children with no disorder or with other disorders, correlated significantly with parent and teacher ratings of hyperactivity and other lab measures of impulsivity, and had adequate test–retest reliability (Gordon & Merttman, 1988). However, others (Barkley et al., 1988) have found the task to be insensitive to stimulant drug effects in children with ADHD, and to correlate poorly if at all with ratings of hyperactivity by parents and teachers (Barkley, 1991). One benefit of such measures is that it allows a clinician to observe a youngster in a situation requiring inhibition and sustained attention. But the test has not been examined for its classification accuracy with regard to discriminating either ADHD from no disorder or ADHD from other psychiatric disorders, and so its role in a diagnostic evaluation remains to be established.

Measures of Activity

The measurement of activity level was discussed in a prior edition of this chapter (Barkley, 1990). Given that there have been no clinically meaningful advances in this field of study since that time, no further comment about such measures is made here. Although advances have been made in various technologies for the measurement of activity levels in children with ADHD (see Matier-Sharma et al., 1995; Teicher, Ito, Glod, & Barber, 1996), such improvements are more useful in conducting research investigations of activity level than in making clinically accurate diagnoses. For instance, Matier-Sharma et al. (1995) examined the classification accuracy of a solid-state activity recording device (the actigraph) in judging children as having either ADHD or no disability and then as having either ADHD or another psychiatric disorder. In the first comparison, the activity measure accurately classified 91% of children with high activity scores as having ADHD. But it accurately classified children with low activity scores as having no disability at a level of just 36%. The situation for the comparison of ADHD to other psychiatric disorders was worse. The presence of a high activity level predicted the presence of ADHD in this case with only 77% accuracy; the presence of an average score predicted the presence of a non-ADHD disorder with an accuracy of just 63%. Such figures do not support a recommendation to use the activity measurement in the differential diagnosis of children with ADHD from either nondisabled children or children with other disorders. A parent and teacher rating scale of hyperactive–impulsive behavior would be a more economical and ecologically valid means of assessing this dimension of behavior than would the use of laboratory activity-recording devices.

Projective Measures

No published studies demonstrate the predictive validity of projective tests (such as drawings, inkblots, or storytelling techniques) for the identification of ADHD. However, there is some indication of differences between children with ADHD and nondisabled children on the Rorschach inkblots (Bartell & Solanto, 1995; Gordon & Oshman, 1981; Cotungo, 1995). Indices of impulsivity on the Rorschach may correlate with an objective measure of impulsivity (Ebner & Hynan, 1994); however, this does not necessarily make the Rorschach useful in evaluating ADHD, as no evidence is available for the positive and negative predictive power of this test for ADHD, especially relative to other clinical disorders. Some evidence is also available that children with ADHD may differ from nondisabled children on the Thematic Apperception Test (Costantino, Collin-Malgady, Malgady, & Perez, 1991). But the data do not support use of this test, either, for judging whether a child suffers from this disorder. The gist of these studies is that children with ADHD are indeed more impulsive and easily frustrated than controls. Groups with ADHD also demonstrate more intense feelings of loneliness and dependence, and come across as more avoidant and socially uncomfortable. Yet much of this information could have been more economically obtained, and probably with greater ecological validity, from the broad-band parent and teacher rating scales and/or from the parent and teacher clinical interviews discussed in Chapter 8 (this volume).

Projective measures might be useful when questions are raised about the possibility of serious thought or emotional disturbance. In our opinion, most of the best diagnostic indicators for ADHD are only valid for children who do not otherwise display the more devastating