Journal of Clinical Child Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hcap19

Upgrading the Science and Technology of Assessment and Diagnosis: Laboratory and Clinic-Based Assessment of Children With ADHD

Mark D. Rapport, Kyong-Mee Chung, Gail Shore, Colin B. Denney & Patti Isaacs
Version of record first published: 07 Jun 2010.

Journal of Clinical Child Psychology
2000, Vol. 29, No. 4, 555–568

Upgrading the Science and Technology of Assessment and Diagnosis: Laboratory and Clinic-Based Assessment of Children With ADHD

Mark D. Rapport
Department of Psychology, University of Central Florida

Kyong-Mee Chung, Gail Shore, Colin B. Denney, and Patti Isaacs
Department of Psychology, University of Hawaii

Reviews the usefulness of clinic-based and laboratory-based instruments and paradigms for diagnosing attention deficit hyperactivity disorder (ADHD) and monitoring treatment effects. Extant literature examining the performance of normal children and those with ADHD on an extensive range of neurocognitive tests, tasks, and experimental paradigms indicates that particular types of instruments may be more reliable than others with respect to detecting between-group differences. We review task parameters that may distinguish the more reliable from less reliable instruments. The value of clinic-based and laboratory-based instruments for monitoring treatment response in children with ADHD is questionable when evaluated in the context of ecologically relevant variables such as classroom behavior and academic functioning. We present a general conceptual model to highlight conceptual issues relevant to designing clinic-based and laboratory-based instruments for the purposes of diagnosing and monitoring treatment effects in children with ADHD. Application of the model to currently conceptualized core variables indicates that attention and impulsivity–hyperactivity may represent correlative rather than core features of the disorder. We discuss implications of these findings for designing the next generation of clinic-based and laboratory-based instruments.

Consummating a diagnosis of attention deficit hyperactivity disorder (ADHD) in children is a multifaceted, time-consuming endeavor that is complicated by multiple factors. Contemporary approaches entail the use of clinical interview, extensive history taking (developmental, educational, psychiatric, family, and psychosocial), broadband and narrowband parent–teacher rating scales, review of psychoeducational test data, and direct observation (Rapport, 1993). Shortcomings associated with standard clinical approaches have been reviewed (see Frick, this issue) and suggest an increasing need to consider alternative or complementary strategies for determining diagnostic standing and assessing treatment outcome. Clinic-based and laboratory-based instruments may serve this purpose, as well as more heuristic functions that promote consideration of conceptual models of ADHD.

Designing clinic-based and laboratory-based assessment paradigms for the purposes of diagnosing ADHD and measuring treatment outcome may yield several benefits to the field of clinical child psychology. Chief among them is their potential cost benefits in light of the time required to complete an extensive clinical evaluation of children suspected to have ADHD. Even if the design of highly reliable instruments cannot be immediately realized, they may play a complementary role in improving the sensitivity and specificity of diagnostic decisions while providing information concerning individual differences in treatment response. From a heuristic perspective, well-designed instruments can serve the dual purpose of furthering a more thorough conceptualization of ADHD and providing the means by which to manipulate discrete independent variables to determine their relation with hypothesized functional mechanisms and behaviors of interest.

The central thesis of this article focuses on key issues concerning the development of clinic-based and laboratory-based assessment instruments. We begin with a comprehensive review of past neuropsychological tests, tasks, and experimental paradigms reported in the literature and evaluate their efficiency in differentiating children with ADHD from normal children. We compare characteristics of reliable and unreliable instruments to discern whether they differ with respect to particular dimensions or performance controlling task parameters. The ensuing section appraises

Requests for reprints should be sent to Mark D. Rapport, Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816–1390.

the sensitivity of instruments used to assess treatment response in children with ADHD and evaluates the degree to which they mirror corresponding changes in classroom settings. We subsequently describe a conceptual model to elucidate expected relations among core and peripheral features of ADHD. Complementary analysis of predicted relations stemming from the model provides a preliminary examination of whether attention and impulsivity–hyperactivity are best viewed as core or correlative features of the disorder. We emphasize implications of these findings for designing diagnostically sensitive instruments and follow with a discussion of desirable psychometric properties and conceptual issues.

Neuropsychological Tests, Tasks, and Experimental Paradigms Used to Differentiate Children With ADHD From Their Peers

We located a total of 142 studies published between 1980 and 1999 that compared the performance of children with ADHD and normal controls on neurocognitive tests, tasks, and experimental paradigms by means of computer search and back-search methodology (see Table 1). We excluded from the review studies (a) that failed to include sufficient statistical information regarding between-group differences and (b) in which the participant pool overlapped with participants participating in other published studies by the same research team (the most recently published study was included in these cases). Reviewed studies reported a total of 439 ADHD–normal comparisons using instruments listed in the table (most studies reported between-group comparisons for more than one instrument). We used names of instruments as opposed to construct headings in Table 1 because of a paucity of established construct validity for nearly all tasks reviewed and disagreement in the field concerning the assignment of tasks to specific (e.g., impulsivity) and broad (e.g., executive function) construct domains. Instruments are listed alphabetically using the formal name or description of the task employed, with the number reporting statistically significant between-group differences (ADHD vs. normal children) and percentage of significant findings shown in the corresponding columns of the table. Summary totals and averages (percentages reporting significant between-group differences) by category grouping and for all studies reviewed are also shown in the table (see Total rows in Table 1). A miscellaneous heading is included in the last part of the table and coupled with a brief description (e.g., attention related, language related) to characterize experimental tasks without formal names.

Prior to reviewing the results, several caveats are in order. First, estimates shown in Table 1 may be inflated because studies reporting significant between-group differences for any of their reported outcome variables on a particular task (e.g., omission errors [OEs]) were counted as significant. Second, studies incorporating identical task parameters for a particular instrument, such as the Continuous Performance Test (CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956), were rare despite invoking a common name to describe the paradigm. The extent to which differences in task-controlling parameters affects group summary statistics is currently unknown. Finally, basic psychometric properties (e.g., test–retest reliability) were infrequently reported for a majority of the tasks shown in the table.

Scrutiny of Table 1 illuminates two relevant issues. First, a plethora of instruments have been employed in an attempt to differentiate children with ADHD from normal controls with little emphasis on conceptual or theoretical analysis of the constructs underlying them. Second, performance on nearly all measures requires comparative judgments of similarities or differences either between pairs of stimuli or between concretely presented objects (e.g., letters, numbers, visual figures) and cognitive representations of target stimuli. These judgments culminate in classification of stimuli as either targets or nontargets and are operationally measured in terms of either OEs or commission errors (CEs). For example, stop signal and go/no-go tasks require children to terminate repetitive behaviors in response to particular signals. Although the mathematical model underlying these tasks emphasizes measurement of response time to termination of a repetitive behavior rather than errors, the inference that the decision process at work involves a comparison of visually presented stimuli to a cognitive representation of the stop signal is inescapable. Moreover, reaction time to termination of a behavior is only a continuous (vs. categorical) expression of a CE. Parallel arguments can be made for the majority of instruments employed in the assessment and treatment of children with ADHD.

Potentially reliable and unreliable instruments are shown in Table 2 to permit appraisal of task-controlling parameters that may differentiate between the two groupings. Reliable instruments were those studied by at least three independent investigators that yielded statistically significant between-group differences in at least 75% of the studies. Unreliable instruments showed significant between-group differences in 25% or fewer independent studies.

Central features and characteristics of reviewed tests, tasks, and experimental paradigms used to differentiate children with ADHD from normal controls are highlighted in the table to facilitate comparisons between reliable and unreliable instruments. Most reliable tasks rely on recognition, recall, or some

Table 1. *Neurocognitive Instruments, Tests, or Tasks Used to Differentiate Children With ADHD From Normal Controls*

Task Name or Description	Total Studies	Significant Numbers (*n*)	Significance Unknown (*n*)	Findings (%)
Boston Naming	3	1		33
Brown Peterson[a]	2	2		100
Cancellation				
Digits[a]	3	2		67
Letters	6	4		67
Shapes	6	4		67
Total	15	10		67
CPT				
CPT, Auditory	7	3		43
CPT, Visual				
X Only	21	18		86
AX	30	28		93
BX	3	1		33
Combined	8	8		100
Total	69	58		84
COWAT	4	2		50
CVLT	3	2		67
Door Opening[a]	3	1		33
Finger Tapping	1	0		0
Go/No-Go[a]	3	3		100
Mazes	13	8	1	67
MFFT	18	12	1	71
Paired Associate Learning	4	2	1	67
Pegboard	6	1		17
Rapid Naming	5	2		40
Rey AVLT	3	1		33
ROCF	11	6		58
Stop Signal	9	7		78
Stroop Color Word	17	11	1	69
Tower				
Tower of Hanoi	3	2		67
Tower of London	1	0		0
Total	4	2		50
Trail Making				
Trails A	7	1		14
Trails B	9	4	1	50
Total	16	5	1	33
WCST	18	10		56
WISC–R				
Verbal Subtests				
Arithmetic	5	3	1	75
Coding	4	3		75
Comprehension	2	1		50
Information	2	2		100
Similarities	3	1		33
Vocabulary	12	4		33
Performance Subtests				
Block Design	12	5		42
Digit Span	9	6		67
Mazes	3	0		0
Object Assembly	2	0		0
Picture Arrangement	2	1	1	100
Picture Completion	2	0		0
Total	58	26	2	46
WRAML[a]	3	1		33

(Continued)

Table 1. *(Continued)*

Task Name or Description	Total Studies	Significant Numbers (*n*)	Significance Unknown (*n*)	Findings (%)
Miscellaneous Tasks				
Attention Related	37	22	2	63
Classification	4	2	1	67
Digits	10	6	1	67
Inhibition	5	3		60
Language Related	4	0		0
Matching	7	3		43
Motor Related	17	9		53
Reward	5	3		60
Trail Making	3	1		33
Verbal Fluency	10	4		40
Verbal Memory	15	9	1	64
Visual Motor Sequencing[a]	4	3		75
Visual Recognition	5	2	1	50
Visual Recall	15	11	1	79
Visual Search	10	6		60
Total	151	84	7	58
Totals	439	257	17	61
Total Studies Reviewed	142			

Note: ADHD = attention deficit hyperactivity disorder; CPT = Continuous Performance Test; COWAT = Controlled Oral Word Association Test; CVLT = California Verbal Learning Test; MFFT = Matching Familiar Figures Test; AVLT = Auditory Verbal Learning Test; ROCF = Rey-Osterrieth Complex Figure; WCST = Wisconsin Card Sorting Test; WISC–R = Wechsler Intelligence Scale for Children–Revised; WRAML = Wide Range Assessment of Memory and Learning; X = CPT version in which the letter *X* is the designated target; AX = CPT version in which the letter *A* followed by the letter *X* is the designated target; BX = double letter version of the CPT in which any letter that immediately repeats itself is the designated target.

[a]Less than three independent investigations.

Table 2. *Characteristics of Reliable and Unreliable Tests, Tasks, and Paradigms Used to Differentiate Children With ADHD From Normal Controls*

				Working Memory				
Task	Recognition	Recall	Time Parameter	Subvocal Speech	Buffer	Response Stimulus Present	Pace	Effect Size
Reliable Tasks								
CPT	Yes	No	Seconds or milliseconds	Yes	Yes	No	O	0.85
Stop Signal	Yes	No	Milliseconds	Yes	Yes	No	O	1.03
WISC–R Arithmetic	No	Yes	Seconds	Yes	Yes	No	S–O	0.92
WISC–R Coding	Yes	Yes	Seconds to minutes	Yes	Yes	Yes or No	S–O	0.78
Visual Recall (Miscellaneous)	No	Yes	Seconds	Yes	Yes	No	S/O	0.78
Unreliable Tasks								
Finger Tapping	No	No	Seconds to minutes	No	No	Yes	S	0.27
WISC–R Mazes	No	No	Seconds to minutes	No	No	Yes	S	0.34
Pegboard	Yes	No	Seconds to minutes	No	No	Yes	S	0.33
Trails A	Yes	No	Minutes	No	No	Yes	S	0.37
Language (Miscellaneous)	No	Yes	Minutes	No	No	No	S	0.47

Note: ADHD = attention deficit hyperactivity disorder; CPT = Continuous Performance Test; S = self; O = other;WISC–R = Wechsler Intelligence Scale for Children–Revised.

combination of the two. Recognition tasks provide children with a copy of the information they need to find in memory. That is, one compares the representation of the perceptual input with representations stored in memory. Conversely, recall tasks require children to access representations different from that of the question. Reliable tasks tend to involve more speeded processing than do unreliable tasks (i.e., milliseconds to seconds vs. seconds to minutes, respectively) and place special demands on working memory. For example, reliable instruments place clear demands on the subvocal speech and buffer components of the articulatory control loop, whereas unreliable tasks do not. The articulatory control, or phonological loop, is viewed as a central component of working memory and is used for storing verbal information or material or nonverbal materials that can be translated into verbal code. Phonologically based information is automatically rehearsed by means of the subvocal speech mechanism (i.e., material to be remembered is pronounced without sound, or subvocally) and creates a record in a phonological buffer that rapidly fades unless rerehearsed immediately (for a thorough discussion, see Baddeley, 1998). Each of the reliable tasks listed in Table 2 uses stimuli that must be encoded phonologically, rehearsed subvocally, and held in working memory for successful execution.

A fourth distinction between the reliable and unreliable instruments concerns whether the response stimulus is present throughout the test trial (i.e., the answer or stimulus required to generate a correct response is available to the child throughout the trial). As shown in the table, none of the reliable instruments display the response stimulus continuously throughout test trials, whereas most unreliable instruments do. The only exception to this pattern is Wechsler Intelligence Scale for Children–Revised (Wechsler, 1974) coding, in which stimuli are present throughout the task, but children must alternate their attention between printed codes to be remembered and a response sheet. A final characteristic that may differentiate reliable from unreliable instruments is the pacing dimension. All of the reliable instruments are to some extent experimenter paced (marked by an *O* in Table 2), in which children have minimal or no control over the presentation rate of stimuli or are under strict timing criteria. In contrast, unreliable instruments nearly always involve self-pacing, whereby children have greater control over the rate by which stimuli are presented.

Average effect sizes for both reliable and unreliable tasks are shown in the last column of Table 2 and indicate the magnitude of a particular type of task for differentiating performance between children with ADHD and normal children expressed in standard deviation metrics. Effect size calculations were based on recommended formulas (Hedges, 1982) and involved subtracting treatment means from control group means and dividing by the pooled standard deviation of the two groups. Larger effect sizes are desirable and indicate greater mean differences in performance between children with ADHD and normal controls.

Several tentative conclusions and caveats concerning the use of tests, tasks, and experimental paradigms for the purposes of diagnosing and understanding core features in children with ADHD are suggested by the foregoing review. The patterning of results imply that performance-controlling task parameter manipulations guided by theory pertaining to comparative judgments and speeded recognition processes are likely to yield insights into etiologically relevant and treatment-related cognitive processes implicated in ADHD. A rich body of literature exists pertaining to these processes and their mediation by working memory, but these processes have typically been viewed as correlates of ADHD rather than a set of core causal cognitive processes.

Other than its potential heuristic value, the foregoing review will no doubt prove disappointing because knowing about significant between-group differences are of no practical consequence to clinicians. The clinical usefulness of clinic-based instruments rests on their ability to provide unique or incremental information to the diagnostic process. Information concerning the sensitivity and specificity of a given instrument and whether it provides incremental diagnostic validity in combination with nonexperimental procedures (e.g., structured interviews, semistructured interviews, rating scales) is particularly relevant. Extant research examining the usefulness of commonly used clinic-based instruments for differentiating between children with ADHD and other childhood disorders is scarce and suggests a clear lack of diagnostic specificity (Werry, Elkind, & Reeves, 1987). Others have reached similar conclusions concerning the ecological validity of laboratory and clinic-based instruments (Barkley, 1991). These findings should bear little surprise and are not entirely disheartening. The review highlights the clear dearth of a conceptual approach to understanding and assessing ADHD. Discovery of useful instruments for diagnosing ADHD will not prove to be a particularly difficult enterprise but must be preceded by experimental investigations and manipulations that enucleate core mechanisms and processes inherent to ADHD.

Despite finding that none of the reviewed tasks, tests, or experimental paradigms are currently appropriate for diagnosing ADHD does not distract from their potential value for assessing treatment outcome. Assessment of treatment outcome, albeit also dependent on established reliability, requires a somewhat different set of desirable psychometric properties, such as the sensitivity to detect clinically meaningful change in children's behavior as a function of receiving treatment or as a function of different levels of treatment. We examine this issue in the ensuing section.

Sensitivity of Instruments for Detecting Treatment Effects in Children With ADHD

The use of tests, tasks, and experimental paradigms for purposes of assessing treatment effects in children with ADHD is limited almost exclusively to studies of drug effects. The lone exception is an intervention study employing cognitive behavior therapy, in which the Matching Familiar Figures Test (MFFT; Kagan, Rosman, Day, Albert, & Phillips, 1964) was used as an outcome measure and failed to detect treatment effects (Fehlings, Roberts, Humphries, & Dawe, 1991). Extensive reviews are available concerning the effects of psychostimulants and other drugs on cognitive function in children with ADHD (for reviews, see Denney & Rapport, 2000; Rapport & Denney, 1999; Rapport & Kelly, 1991; Solanto, 1998). The central question regarding treatment sensitivity, however, is not whether drugs affect cognitive performance but whether clinic-based and laboratory-based instruments are incrementally useful for evaluating and adjusting medication dosage to help children function in their everyday environments. Thus, the ensuing review focuses exclusively on comparing children's performance on popularly used clinic-based and laboratory-based instruments with corresponding measurement of their classroom functioning by the use of standardized teacher ratings.

Experimental Design and Medication Protocol

For illustrative purposes, the results obtained from clinic-based assessment and classroom observation for 36 children with ADHD are presented here. The sample represents a smaller subset of children who were evaluated using each of the instruments described in the following paragraphs (only a subset of children culled from the larger sample were administered all tests). Detailed description of diagnostic criteria, methods, and clinical outcome for the larger sample have been reported (see Rapport, Denney, DuPaul, & Gardner, 1994). Briefly, all children met stringent diagnostic criteria for ADHD according to semistructured clinical interview and parent–teacher ratings on standardized instruments. A double-blind, placebo-controlled, within-subjects design (crossover) experimental design was used in which children received a placebo and each of four methylphenidate (MPH) doses after baseline assessment. The order of dose administration was counterbalanced and determined by random assignment such that an equal number of children received each dose during a given week of the study. MPH was prescribed by each child's physician in the following doses: placebo, 5 mg, 10 mg, 15 mg, and 20 mg. After baseline data collection (1st week), parents were given 1 week of medication in predated envelopes at a single-dose level. This procedure continued until each child received every dose for 6 consecutive days. All weekly dose changes occurred on Sundays (no capsules were administered on Saturday) to allow for "washout" and to control for possible rebound effects.

Clinic Assessment

Children were tested once weekly at the Children's Learning Clinic with standardized procedures under baseline and each of the aforementioned counterbalanced drug conditions. Testing began 60 to 90 min after oral ingestion because of the short behavioral time response of MPH (Swanson, Kinsbourne, Roberts, & Zucker, 1978). Clinic assessment included standardized administration of the Paired Associate Learning Task (PALT; Rapport, Loo, & Denney, 1995), CPT (Rapport et al., 1987), and MFFT (Rapport et al., 1988) in a counterbalanced sequence. These tasks are the most commonly used clinic-based assessment instruments for assessing overall drug response and between-dose differences in children.

Classroom Assessment

Children were observed in their regular classrooms for 20-min intervals, 3 days per week across the 6-week evaluation period. Observations were completed during morning hours because of the behavioral time-response course of MPH and began 1.5 to 2 hr after children received their morning medication. During each observation period, children completed their usual in-seat academic work assigned by the classroom teacher (e.g., mathematics or language arts assignments). Three dependent variables were obtained and compared with clinic-based assessment measures to address questions concerning the sensitivity of the latter for mirroring changes in children's classroom functioning.

Teacher ratings. Classroom teachers completed the Abbreviated Conners' Teacher Rating Scale (ACTRS; Werry, Sprague, & Cohen, 1975) each Friday throughout the study. The outcome reflected the children's behavior during the morning hours (until 11:30 a.m.) of the preceding week. All teachers were blind to when medication was administered and to the specific doses.

Attention. Directly observed on-task behavior was used as a measure of attention. On-task behavior emphasizes visual fixation to task-relevant stimuli, a property common to most measures of attention, and exhibits superior precision, objectivity, and validity as a measure of classroom attention relative to teacher-rated scales. Trained undergraduate and graduate-level research assistants observed children for 60 consecutive intervals during each observation period (3 days per condition) throughout the study. Each interval was divided into 15 sec of observation followed by 5 sec of recording. A child's behavior was categorized as either on or off task. *Off-task behavior* was defined as visual inattention to one's materials for more than 2 consecutive seconds within each 15-sec observation interval, unless the child was engaged in an alternative task-appropriate behavior (e.g., sharpening a pencil). Observers were situated in classrooms such that they avoided direct eye contact with target children and were distanced from them by approximately half the classroom size, although they were still able to clearly determine task-related attention. Observers were blind to children's medication status throughout the study.

Academic efficiency score (AES). Children's performance on regularly assigned academic seat work during scheduled observation periods was used as a dependent measure to preserve ecological validity yet maintain adequate experimental control. Classroom teachers assigned academic seat work consistent with the children's ability levels but with the stipulation that assignments be worked on during the optimal medication periods (1.5–2 hr after medication) and assignments be gradable in terms of percentage completed and percentage accuracy. Either the teacher or the primary observer graded assignments after class. Percentage completion and percentage accuracy rates were recorded daily. The two scores were combined to yield an AES, defined as the percentage of assigned work completed correctly.

Reliability of observational measures. Interobserver reliability checks of each child's on-task behavior were obtained on 33% of the observation days and at least once per week for each participant in the study. Obtained and chance estimates were computed for occurrence, nonoccurrence, and overall agreement. Overall reliability for the larger group (n = 76) was consistently greater than 85%, with a mean of 92.4% across children. A mean kappa value of .84 was obtained across all observations. Overall reliability for AES was consistently greater than 95%, with a mean of 97% and a kappa coefficient of .96 (see Rapport et al., 1994).

Dose Sensitivity of Clinic-Based Instruments for Mirroring Classroom Functioning in Children With ADHD

A set of miniplots was constructed to illustrate the relation between children's performance on clinic-based measures (CPT, MFFT, PALT) and observations of their functioning in regular classrooms across a wide range of MPH doses. The miniplots show the percentage of possible improvement on each dependent measure from baseline conditions. Values were obtained by determining the range of possible change relative to baseline for each dependent measure in standard (T) score form and calculating the percentage of change associated with each dose condition. Scores were converted in this manner so the measures could be placed on the same scale for comparison purposes and to equate for differences in the degree or range of change possible across measures.

Inspection of Figure 1 indicates that none of the three clinic-based instruments provided an accurate assessment of treatment effects based on direct observations and teacher ratings of children's classroom functioning. Two potential reasons for the apparent insensitivity of the clinic measures include practice or placebo effects and a failure to differentiate between low-, intermediate-, and high-dose conditions. Changes due to placebo or practice effects are shown in the rectangular boxes labeled *placebo response* at the bottom of each of the three miniplots. Greater susceptibility to these effects are evinced as scores progress from a left-to-right position on the abscissa. As shown, PALT trials to criterion and errors evinced approximately 10% and 16% of their total change from baseline performance conditions under placebo (see upper miniplot in Figure 1). CPT OEs and CEs showed an opposite pattern (see center miniplot in Figure 1) and worsened under placebo conditions as indicated by their movement from right to left of the zero line in the bottom left-hand corner of the figure (children's CPT CEs showed greater deterioration of 48% but could not be plotted because of practical considerations in constructing the graph). The undesirable degree of instability inherent to these two instruments is consistent with clinical observations of improved PALT and worsened CPT performance over time under no-medication conditions.

The dose–response curves shown in the three miniplots of Figure 1 also indicate that change in the clinic-based measures from baseline to medication conditions range from approximately 20% to 30%. This finding stands in contrast to the relatively broad range evinced in the three classroom measures of between 50% to 65% and suggests that the clinic measures are insensitive to dose-related changes in behavior, limited by attributes inherent to the tasks themselves (e.g., floor and ceiling effects), or both.

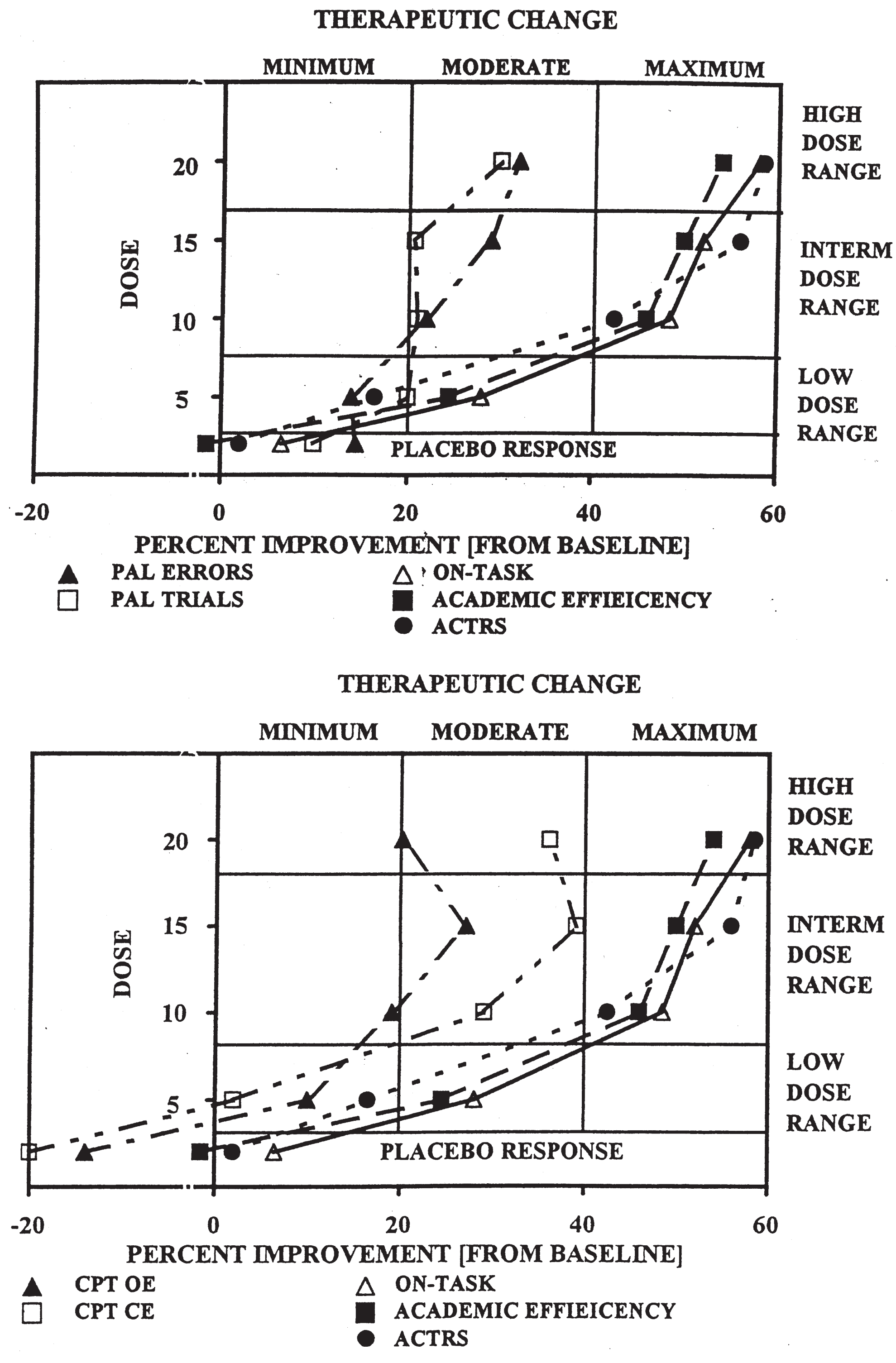

Figure 1. The mean percentage of improvement possible on each dependent measure relative to baseline conditions depicted as dose–response curves and expressed as standard (*T*) scores for the weekly teacher ratings (Abbreviated Conners' Teacher Rating Scale [ACTRS]), classroom observation (on-task and academic efficiency scores), and clinic measures (Paired Associated Learning Task [PAL], Continuous Performance Test [CPT], and Matching Familiar Figures Test [MFFT]). Left-to-right movement on the horizontal axis indicates improvement on all dependent variables. Dose (methylphenidate in mg) is shown on the vertical axis. Therapeutic "windows" highlight dose by change effects. Placebo response is shown in the bottom window of the figure and indicates worsening (movement to the left of the zero line) or improved performance or behavior (movement to the right of the zero line) under the placebo conditions.

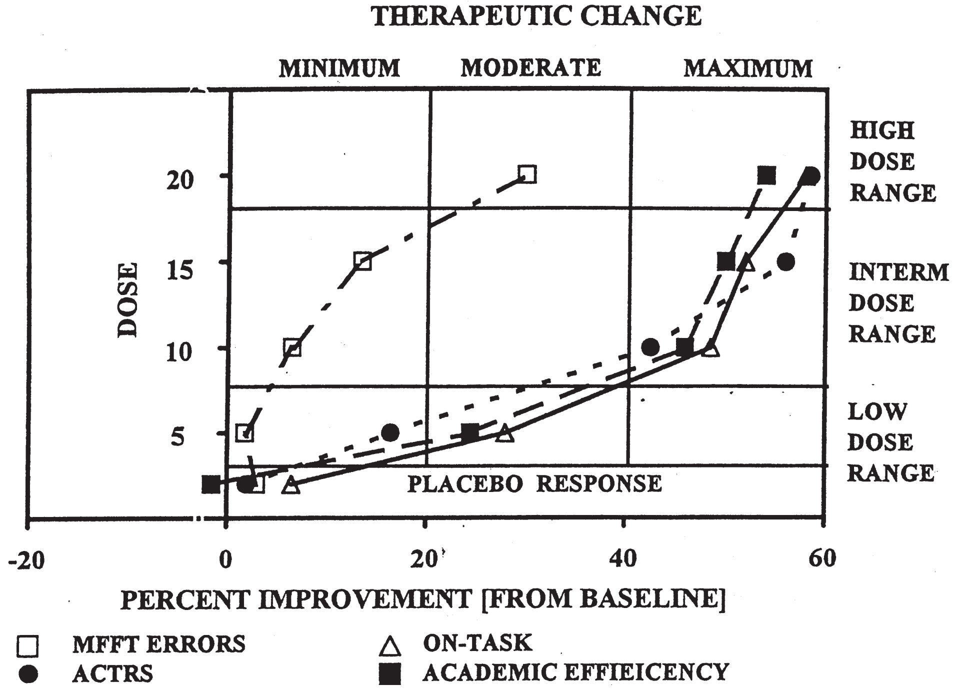

Figure 1. ***(Continued)***

The shape and placement of the curves are also of interest. For example, consider the preponderance of Dose × Change data points falling within a single therapeutic range (minimum or intermediate windows) in Figure 1. Their placement indicates that one difficulty with the clinic-based measures is their failure to differentiate between low- and high-dose conditions, although the two requiring learning (PALT) and problem solving (MFFT) begin to move from a left (minimal change) to right (intermediate change) position on the figure as a function of increasing dose. In contrast, the dose–response curves for CPT OEs and CEs begin to retreat toward baseline levels with increasing dose.

The external validity of the clinic-based instruments reported here for mirroring treatment-related improvement in children's classroom behavior and academic performance remains unproven at present. These findings are consistent with published studies (Rapport et al., 1987; Rapport, Loo, & Denney, 1995; Rapport et al., 1988) and reviews (Barkley, 1991) concerning the ecological validity of clinic-based instruments. As such, no instrument described here can be recommended, alone or in combination, for the expressed purpose of clinically titrating MPH in children with ADHD.

The failure to devise clinically useful tests, tasks, and experimental paradigms for diagnostic and treatment monitoring purposes with respect to ADHD suggests that basic assumptions underlying the constructs they were designed to measure warrant scrutiny. In the ensuing section, we review basic assumptions concerning current conceptualizations of ADHD and evaluate several of their premises by means of structural equation modeling. We follow this with a focused discussion that highlights desirable characteristics and properties of clinic-based and laboratory-based instruments.

Theoretical and Conceptual Issues Underlying the Design of Clinic and Laboratory Instruments for Children With ADHD

The appropriate design of clinic-based and laboratory-based instruments for diagnosing children with ADHD and assessing treatment outcome rests on whether underlying assumptions concerning core deficits associated with the disorder are correctly specified. Elucidation of core deficits, central processes, and the means by which they cause ADHD behavior problems inform us about the types of tasks to develop that will enable adequate challenge to suspected systems by careful design and manipulation of task parameters. A model based on extant literature is illustrated in Figure 2 and elucidates the implicit and explicit causal assumptions underlying current conceptualizations of ADHD. It suggests that biological influences (e.g., genetics, prenatal insults) give rise to individual differences in the functional properties of neurobiological systems (e.g., dopaminergic–noradrenergic neurotransmission) that are etiologically responsible for the core psychological (i.e., cog-

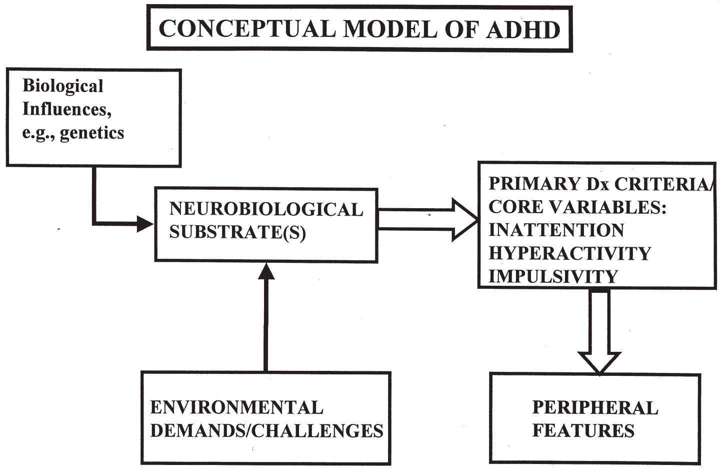

Figure 2. **A conceptual model of attention deficit hyperactivity disorder (ADHD). Biological influences (e.g., genetics) give rise to individual differences in the functional properties of neurobiological systems (e.g., dopaminergic–noradrenergic neurotransmission) that are etiologically responsible for the core psychological (cognitive and behavioral) features of ADHD. Peripheral (secondary) features are conceptualized as causal by-products of core features.**

nitive and behavioral) features of ADHD. Core features of the disorder are currently conceptualized as involving attention and impulsivity–hyperactivity. Peripheral features of ADHD (e.g., those listed as associated features in the *Diagnostic and Statistical Manual of Mental Disorders* [4th ed.]; American Psychiatric Association, 1994) are conceptualized as causal byproducts of the core features. For example, the academic underachievement observed in many children with ADHD is presumed to be a causal consequence of broader, more primary features of the disorder such as chronic inattention to and lack of persistence on classroom tasks. Other peripheral features common to ADHD, such as inadequate social skills and peer relationships, low frustration tolerance, and strained family relationships, are similarly presumed to be due to core behavioral and cognitive influences. The conceptual framework can accommodate either categorical or dimensional views of psychopathology. Thus, neurobiological substrates may be either present or absent and give rise to correspondingly categorical symptomatic presentations, or they may vary incrementally across individual children, giving rise to continuously distributed psychological features. Comprehensive models of the disorder will need to reconcile this empirically observed symptom distribution (Gjone, Stevenson, & Sundet, 1996; Levy, Hay, McStephen, Wood, & Waldman, 1997) with properties of the neurobiological substrates presumed to explain them.

The degree to which attention and impulsivity–hyperactivity accurately represent core deficits of ADHD can be examined empirically by modeling expected patterns of relations predicted by the conceptual model under no-treatment (baseline) and treatment conditions as discussed in the ensuing two sections. Scrutiny of hypothesized core deficits has clear and central relevance to designing performance-based measures of ADHD.

Core Deficits

Attention and impulsivity–hyperactivity are currently viewed as core features of ADHD. The conceptual model implies that measures of core features should demonstrate moderate to high intercorrelations and show expected patterns of relations with peripheral variables to the extent that the latter are causally dependent on the former for successful execution. Structural equation modeling is used to illustrate this conceptual approach and examines expected patterns of relations among core variables and between core and peripheral measures (see Figure 3a). The data and measures were collected as part of an outcome study involving 76 children with ADHD conducted by Rapport and colleagues (Rapport & Denney, 1999; Rapport et al., 1994). Variables shown on the left side of Figure 3a represent baseline observations or ratings for three classroom measures: (a) observed attention to task, (b) teacher ratings of self-control (Teacher Self-Control Rating Scale; Humphrey, 1982), and (c) hyperactivity (ACTRS; Werry et al., 1975). The outcome variable on the right side of Figure 3a represents children's academic performance during baseline (no-treatment) conditions. Classroom academic performance was selected as the peripheral variable because of the extensive literature documenting its short-term and long-term relations with ADHD (for a review, see Rapport, Scanlan, & Denney, 1999).

Intercorrelations among hypothesized core variables are shown by the coefficients adjacent to the dou-

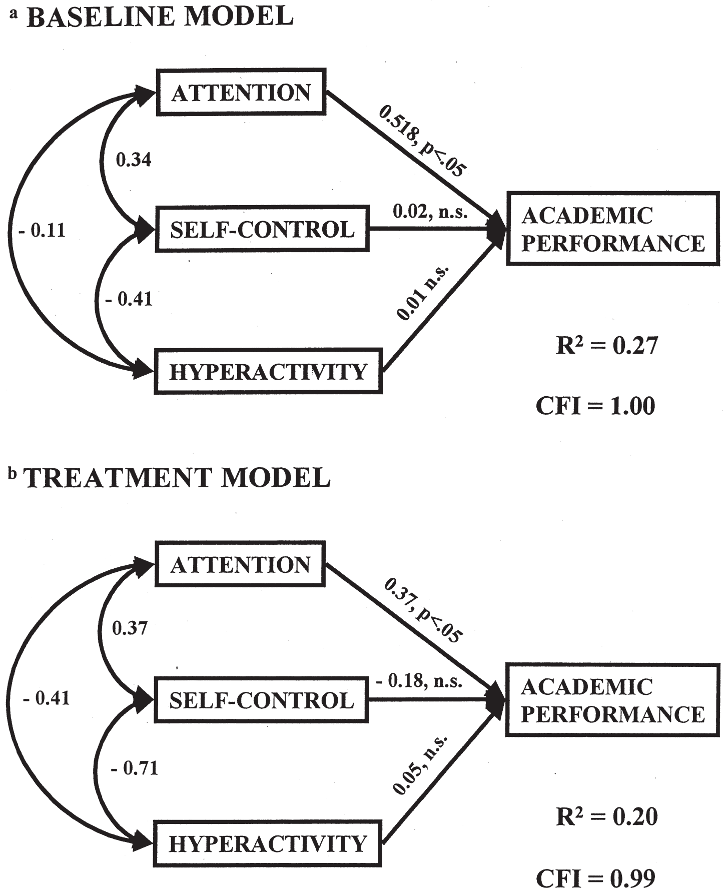

Figure 3. Structural equation models of expected patterns of relations among core features of attention deficit hyperactivity disorder and peripheral variables under baseline (a) and treatment (b; methylphenidate) conditions. Intercorrelations among hypothesized core variables are shown by the coefficients adjacent to the double-headed arrows. Coefficients on the single-headed arrows pointing toward the academic performance variable represent correlations (regression coefficients) between academic performance and each of the three predictors after accounting for intercorrelations among the later. R^2 = percentage of variance accounted for in the model; CFI = comparative fit index.

ble-headed arrows in Figure 3 (i.e., .34, –.11, and –.41). These findings indicate that the ability to stay on task (attentiveness) is positively related to teacher ratings of self-control, with both attention and self-control negatively related to hyperactivity. Collectively, these results are consistent with model predictions in demonstrating that core variables show expected patterns of relations with one another.

Coefficients on the single-headed arrows pointing toward the academic performance variable represent correlations between academic performance and each of the three predictors after accounting for intercorrelations among the latter. These coefficients indicate that attention is related to academic functioning (.52, $p < .05$), whereas teacher rated self-control and hyperactivity are not. Despite a strong model fit (comparative fit index [CFI] = 1.00), the R^2 value beneath the academic performance variable indicates that the core domains of attention, self-control, and hyperactivity account for only 27% of the variation in the peripheral variable, academic performance.

Treatment-Related Changes in Hypothesized Core Domains and Peripheral Variables

Treatment-related improvement in core domains should show strong correspondence with gains in peripheral areas of functioning to the extent that the latter are causal byproducts of the former and rely on core areas for successful execution. A structural equation model consisting of hypothesized core domains (attention, impulsivity, hyperactivity) and a peripheral measure (classroom academic performance) is illustrated in Figure 3b to examine predicted patterns of relations stemming from the general conceptual model. Variables shown on the left side of the figure represent slopes of dose–response curves (based on placebo, 5-mg, 10-mg, 15-mg, and 20-mg doses of MPH) for the same three classroom measures used to examine core deficits mentioned previously. The outcome variable to the right represents dose–response curve slopes of children's weekly academic performance (percentage of academic assignments completed correctly). Intercorrelations among these slopes are shown by the coefficients adjacent to the double-headed arrows (i.e., .37, –.71, and –.41) and confirm expected patterns of relations among core domains. Coefficients on the single-headed arrows pointing toward the academic performance variable represent correlations between rates of change in academic performance and each of the three predictors after accounting for intercorrelations among the latter. These coefficients indicate that gains in attention were related to improved academic functioning (.37, $p < .05$), whereas teacher-rated self-control and hyperactivity were not. Despite a strong model fit (CFI = 0.99), the R^2 value beneath the academic performance variable indicates that the combination of gains in hypothesized core domains of attention, self-control, and hyperactivity account for only 20% of the variation in the peripheral variable academic improvement.

Several explanations may account for this pattern of findings. We briefly review these to highlight the complexity of conceptual issues involved in identifying core deficits. First, core and peripheral variables were estimated in the model using single measures of construct domains. Because of the inherent measurement error and relative impurity of any single measure, incorporating several measures of attention, impulsivity–hyperactivity, and academic functioning with demonstrated high internal consistency would be advisable to provide more accurate assessment of core and peripheral constructs. Second, mediator variables or pathways may be required to correctly model the relations or relevant processes bridging core (attention, impulsivity–hyperactivity) and peripheral variables. For example, Rapport et al. (1999) demonstrated that the relation between attention and long-term scholastic underachievement in ADHD is mediated by working memory processes. Finally, the relatively high intercorrelations among core variables raises the possibility that they may represent complementary operations and behaviors underlying a more central, higher order psychological process.

In summary, modeled relations based on direct observations and teacher ratings of children's behavior under baseline and treatment conditions highlight the potential value of generating empirically derived conceptual models of ADHD prior to designing diagnostically relevant and treatment-sensitive clinic-based and laboratory-based instruments. If inattention, for example, proves to be a core feature of ADHD, increased emphasis should be placed on identifying the source, process, or mechanisms by which attention fails and the conditions under which failed attention occurs. Designing instruments that challenge identified underlying processes and mechanisms related to these deficits should subsequently prove more successful than past efforts. We discuss related conceptual and pragmatic issues concerning the design of clinic-based and laboratory-based instruments in the ensuing section.

Desirable Characteristics of Clinic-Based and Laboratory-Based Instruments

Designing clinic-based and laboratory-based instruments for purposes of diagnosing ADHD and measuring treatment outcome requires examination of several conceptual issues in addition to traditional psychometric considerations. Hypothesized core features of the disorder must be carefully scrutinized and empirically validated, and their expected patterns of relations with other core domains and peripheral variables must be established to elucidate the psychological processes through which these variables are related. Empirical validation of core processes can proceed at multiple levels but eventually must entail careful manipulation of discrete independent variables and observation of their effects to provide compelling evidence of core deficits, their underlying processes, and how these translate into behaviors characteristic of ADHD.

Extant research concerning differentiation of ADHD from other disorders has relied almost exclusively on tests and tasks borrowed from existing neuropsychological or specialized batteries. These are typically designed for more global purposes and provide interesting correlational data (e.g., higher error scores are related to a categorical diagnosis). They fail, however, to provide essential information concerning why particular types of errors are made, underlying processes, and mechanisms by which these processes operate. For example, it is not possible to discern whether OE, CE, or total error scores that exceed normative standards are due to perceptual difficulties, attentional problems, deficient memory processes, or an interaction involving these or other processes. Error scores can be correlated with direct observations to provide some evidence of relations (e.g., OEs and teacher ratings), but a more convincing demonstration could be made by manipulating specific independent variables and observing whether expected patterns of behavior emerge as a result of these manipulations. Explication of performance-controlling parameters and their relations to core and peripheral features represents an important step in this process.

Understanding the influence and potential interactions of task-controlling parameters permits manipulation of conceptually relevant independent variables and concomitant assessment of their effects on expected levels of performance and behavior. For example, if hyperactivity is a byproduct of more central features of ADHD such as attention deficits, it should be possible to challenge the underlying system by manipulating variables that affect or control attentional processes with resulting increases and decreases in gross motor behavior.

General Features of Desirable Instruments

If one assumes that individual differences associated with underlying affected substrates are always present and relatively constant, these differences should be expressed on outcome measures designed to assess core psychological (cognitive and behavioral) features of the disorder and should vary both within and across children depending on demands imposed on the underlying substrate. Creative design of clinic-based and laboratory-based paradigms will succeed to the extent that tasks are designed to impose demands on suspected systems and concomitantly permit assessment of children's tolerance for challenge at both individual and normative levels of functioning.

Variation of scores within and across diagnostic groups and the general population on paradigms designed to assess ADHD is also desirable and need be in accord with theoretical expectations. Instruments designed to permit a wide distribution of scores (vs. the use of artificial cutoffs) are preferred to permit consideration of both categorical and dimensional views of ADHD. Conversely, minimal intra-individual and interindividual variation on outcome measures precludes ascertainment of reliable diagnostic judgments.

Psychometric Features of Desirable Instruments

The establishment of sound psychometric properties incorporating validity and reliability metrics represents the sine qua non of clinical assessment and will provide relief to a domain largely bereft of such standards. Devising paradigms devoid of floor, ceiling, and practice effects will prove particularly beneficial for both diagnostic and treatment monitoring purposes. For this latter purpose, establishing externally valid measures that reflect or mirror children's ability to function in school, at home, and with peers will prove particularly valuable.

Summary and Concluding Comments

Extant literature examining the performance of normal children and those with ADHD on an extensive range of neurocognitive tests, tasks, and experimental paradigms indicates that particular types of instruments may be more reliable than others with respect to detecting between-group differences. We identified task parameters that appear to distinguish the more reliable from less reliable instruments. Information concerning the sensitivity, specificity, and incremental validity of reviewed instruments was sparse and suggests that none can be used with confidence for purposes of diagnostic decision making alone or in combination with standard clinical procedures. The small subset of instruments used to monitor treatment response in children was found to be wholly inadequate for mirroring behavior and academic functioning in classroom settings. A general conceptual model was presented subsequently to highlight conceptual issues relevant to designing clinic-based and laboratory-based instruments for purposes of diagnosing and monitoring treatment effects in children with ADHD. The underlying premise being that improved understanding of core features of ADHD will facilitate the design of clinically useful instruments. Application of the model to currently conceptualized core variables raises the possibility that attention and impulsivity–hyperactivity may represent correlative rather than core features of the disorder. We discussed implications of these findings for designing the next generation of clinic-based and laboratory-based instruments and reviewed desirable psychometric properties.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baddeley, A. (1998). *Human memory: Theory and practice.* Boston: Allyn & Bacon.

Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology, 19,* 149–178.

Denney, C. B., & Rapport, M. D. (2000). Cognitive pharmacology of stimulant drugs in AD/HD. In M. V. Solanto, A. F. T. Arnsten, & F. X. Castellanos (Eds.), Stimulant drugs and ADHD: Basic and clinical neuroscience (pp. 283–302). New York: Oxford University Press.

Fehlings, D. L., Roberts, W., Humphries, T., & Dawe, G. (1991). Attention deficit hyperactivity disorder: Does cognitive behavioral therapy improve home behavior? *Journal of Developmental and Behavioral Pediatrics, 12,* 223–228.

Gjone, H., Stevenson, J., & Sundet, J. M. (1996). Genetic influences on parent reported attention related problems in a Norwegian general population twin sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 35,* 588–596.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92,* 490–499.

Humphrey, L. L. (1982). Children's and teachers' perspectives on children's self-control: The development of two rating scales. *Journal of Consulting and Clinical Psychology, 5,* 624–633.

Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs, 78,* 1.

Levy, F., Hay, D. A., McStephen, M., Wood, C., & Waldman, I. (1997). Attention deficit hyperactivity disorder: A category or continuum? Genetic analysis of a large twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 737–744.

Rapport, M. D. (1993). Attention deficit hyperactivity disorder. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment* (pp. 269–291). Boston: Allyn & Bacon.

Rapport, M. D., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and methylphenidate: Assessment and prediction of clinical response. In L. L. Greenhill & B. B. Osman (Eds.), *Ritalin: Theory and patient management* (pp. 45–69). New York: Mary Ann Liebert, Inc.

Rapport, M. D., Denney, C. B., DuPaul, G. J., & Gardner, M. J. (1994). Attention deficit disorder and methylphenidate: Normalization rates, clinical effectiveness, and response prediction in 76 children. *Journal of the American Academy of Child & Adolescent Psychiatry, 33,* 882–893.

Rapport, M. D., Jones, J. T., DuPaul, G. J., Kelly, K. L., Gardner, M. J., Tucker, S. B., & Shea, M. S. (1987). Attention deficit disorder and methylphenidate: Group and single-subject analyses of dose effects on attention in clinic and classroom settings. *Journal of Clinical Child Psychology, 16,* 329–338.

Rapport, M. D., & Kelly, K. L. (1991). Psychostimulant effects on learning and cognitive function: Findings and implications for children with attention deficit hyperactivity disorder. *Clinical Psychology Review, 11,* 61–92.

Rapport, M. D., Loo, S., & Denney, C. (1995). The Paired Associate Learning Task: Is it an externally valid instrument for assessing methylphenidate response in children with attention deficit disorder? *Journal of Psychopathology and Behavioral Assessment, 17,* 125–144.

Rapport, M. D., Scanlan, S., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *Journal of Child Psychology and Psychiatry, 40,* 1169–1184.

Rapport, M. D., Stoner, G., DuPaul, G. J., Kelly, K. L., Tucker, S. B., & Schoeler, T. (1988). Attention deficit disorder and methylphenidate: A multilevel analysis of dose–response effects on children's impulsivity across settings. *Journal of the American Academy of Child & Adolescent Psychiatry, 27,* 60–69.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. J., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350.

Solanto, M. V. (1998). Neuropsychopharmacological mechanisms of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. *Behavioural Brain Research, 94,* 127–152.

Swanson, J. M., Kinsbourne, M., Roberts, W., & Zucker, K. (1978). Time-response analysis of the effect of stimulant medication on learning ability of children referred for hyperactivity. *Pediatrics, 61,* 21–29.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* New York: The Psychological Corporation.

Werry, J. S., Elkind, G. S., & Reeves, J. C. (1987). Attention deficit, conduct, oppositional, and anxiety disorders in children: III. Laboratory differences. *Journal of Abnormal Child Psychology, 15,* 409–428.

Werry, J. S., Sprague, R. L., & Cohen, M. N. (1975). Conners Teacher Rating Scale for use in drug studies with children: An empirical study. *Journal of Abnormal Child Psychology, 3,* 217–229.

Manuscript received December 10, 1999
Final revision received March 10, 2000