6

Activity Measurement

Mark D. Rapport
Michael J. Kofler
Carmen Himmerich

Department of Psychology
University of Central Florida
Orlando, Florida

INTRODUCTION

Activity level is the first enduring trait or personality characteristic to develop in humans. Individual differences are apparent by the 28th week of gestation, and developmental trajectory studies suggest that activity level follows a curvilinear relationship between infancy and late adulthood. Its relationship with children's behavioral characteristics is intriguing. Heightened activity level following the neonatal period is associated with desirable behavioral attributes such as positive social interactions, motor and mental maturity, and inquisitiveness. This association rapidly reverses during preschool and early elementary school years. Pejorative characteristics such as aggression, distractibility, academic underachievement, learning disabilities, and strained peer and parent–child relationships are assigned to children exhibiting heightened activity after 5 years of age. These difficulties continue into middle childhood for many children, and they set the stage for a lifetime of disabilities, despite the diminution in activity level typically observed during adolescence.

Studies of activity level in children traditionally focus on understanding four interrelated areas: (a) the relative contribution of genetic, biological, and socialization developmental factors relevant to activity level; (b) the effects of activity level on physical growth and motor skill development; (c) age and gender trends in activity level across settings and situations; and (d) the role of activity level in understanding temperament or enduring personality differences. Collectively, this literature provides a strong foundation for understanding the origin, development, and expression of activity level in children (for a review, see Eaton, 1994). The next chapter section addresses the normal development and heritability of activity level in children as a reference for understanding abnormal activity—particularly age and gender effects.

Measurement of activity level in children is interesting in its own right, and it poses unique challenges to activity level research. It runs the gamut from subjective questionnaires to highly sophisticated and reliable actigraphs costing thousands of dollars. Sophistication and reliability alone, however, may be insufficient if our purpose is to understand the conditions under which activity level differs among children and how best to assess these differences. Consider the actigraph and its highly detailed output. Millisecond changes in activity level may be collected for several weeks at a time and downloaded to a computer to scrutinize differences between children with and without a particular clinical disorder. We determine that one group of children moves more frequently than another during the day. An interesting finding in its own right, but less than satisfactory if we wish to understand the underlying dynamics that contribute to the between-group differences. Combining the actigraph information with a daily activity log reveals that the first group of children is more active than the second group only when engaged in tasks or activities that require them to sit down and work on academic tasks. Our understanding of group differences gains further momentum as we add more information to the raw actigraph data, such as a videotape of the children's movements throughout the day. Video footage is analyzed using a computer scoring system. The analysis confirms between-group differences during academic assignment periods and reveals that activity level in the more active group of children constitutes two functionally distinct types of movement—one that appears to help them maintain alertness to the assignment, and the other, a form of escape behavior that removes them from the task. Our understanding of the phenomenon has increased exponentially from a simple statement of higher motor movement frequency in one group of children relative to another, to an inchoate appreciation of the dynamic interplay among activity level, cognitive function, and behavior. Other researchers may wish to explore the findings from a physiological perspective or use controlled laboratory investigations. The first group seeks to explain possible underlying differences in anatomical structure or processes serving these structures, whereas the second group is interested in determining whether particular cognitive variables or processes contribute to differences in activity level. The move from simple to complex and complementary measures of activity level in children is explored in the subsequent chapter section, which highlights the critical role assessment plays in scientific discovery.

EARLY DEVELOPMENT AND GENDER

A normative-developmental framework is adopted for examining activity level in developing children. This approach lays the groundwork for determining what is atypical and potentially pathological at other stages of development. For example, some instances of activity level may represent normal quantitative variations evident at other developmental periods or in less intense degree or form in same-age children. Other instances may represent qualitative variations that are not normal for any developmental period.

ACTIVITY LEVEL IN INFANCY

Differences in human activity level are evident by the 28th week of gestation and predict several aspects of infant behavior. For example, fetal movement recorded during the last three months of pregnancy is significantly correlated with infants' motor, language, and personal-social skills at 12 weeks of age (Walters, 1965). Children with higher activity levels in utero evidence more advanced motor development at 8–12 weeks, including lifting their head for a prolonged period of time when lying on their stomach, recognizing familiar voices, using variations of cry tones to signify different needs, and smiling responsively to their parents. The cephalocaudal developmental sequence of activity level—beginning with the head and proceeding to the arms and finally to the legs—suggests a strong genetic influence.

Increases in activity level following the neonatal period are readily noticeable in infants and include rhythmical stereotypic movements involving the extremities, such as repetitive hand movements and leg kicks. Desirable behavioral characteristics begin to develop at this stage. Infants become interested in social interactions, recognize pleasant objects such as a bottle, and react with excitement and increased activity to a wide range of environmental stimuli. The corresponding development of locomotor behavior is easily observed in maturing infants. Movement becomes increasingly more coordinated, purposeful, and functional during the first 18 months of life. These rapid changes are evident in the progression of developmental milestones, such as lifting the head (approximately 2–3 months of age), rolling over (4–6 months), sitting unassisted and crawling (7–9 months), standing (10–12 months), and eventually walking unassisted (12–18 months). The sequence of these milestones is typically the same in children, with occasional variations. Activity level peaks at approximately 8 years of age (based on a review of 12 studies involving 840 children between 1 and 24.6 years of age), and these findings are relatively consistent across different cultures (Eaton, McKeen, & Campbell, 2001).

ACTIVITY LEVEL AND HERITABILITY

Heritability plays an important role in determining activity level. Twin studies reveal that a significant degree of infant activity is genetically determined. Monozygotic (identical) twins are significantly more similar in activity level than dizygotic (fraternal) twins at 30 weeks, based on actometer and parent ratings (Saudino & Eaton, 1991). This finding has been confirmed by every published twin study of activity level to date, but it is weakened somewhat because twin study methodology cannot completely control for or rule out third-factor influences. Selective breeding experiments, however, are designed to directly address the degree to which activity level is inheritable. A compelling example is derived from a study involving 30 generations of selective breeding (DeFries, Gervais, & Thomas, 1978). Ten random litters were randomly chosen from 40 possible litters of mice, and the most active and inactive females were selected to become progenitors of a high- and low-activity breeding line. Replicate (high- and low-activity) and control lines were also established. High-active, low-active, and control (intermediate activity) mice were subsequently mated at random within their respective line (e.g., high active with high active, low active with low active) to produce first-generation offspring. The most active male and female were selected from each subsequent litter in the high-active line, and these same procedures were followed for the low-active and control lines for 30 generations of selective breeding. The authors report a tenfold difference in activity level in open field activity between the two highly inbred stains (high and low active), with the control group falling intermediate between the two groups. High-active mice became significantly more active, low-active mice became less active, and intermediate-active mice remained about the same after 30 generations of breeding. These results provide compelling evidence that activity level is significantly influenced by genetic factors.

ACTIVITY LEVEL IN EARLY CHILDHOOD

Motor behavior continues to show an upward trajectory in toddlers and early preschoolers, but the changes are less pronounced and slower relative to those observed during infancy. Development of gross and fine motor behavior dominates this period as children explore and interact with their environment and acquire myriad skills, ranging from using scissors and crayons to riding a tricycle. Environmental and particularly setting effects may significantly influence children's activity level during this time. Some children attend nursery schools and day-care facilities, whereas others have limited access to playgroups and other children. The stability of children's activity level over this time period is remarkable, despite differences in context and environment. For example, the test–retest correlation for a sample of 129 boys and girls assessed at age 3 and again at age 4 was .44 and .43, respectively. This finding indicates strong continuity in children's activity level at a time when development is proceeding rapidly.

The relationship between age and activity level again changes rapidly in late preschool and elementary school (age 5–10), but for the first time it shows a *decline*, due in part to setting effects and societal demands. Children are expected to sit and engage in academic tasks and other cognitive activities for increasingly longer periods of time. Those able to do so are praised for their concentration abilities and tenacity, with accompanying high grades and test scores. Pejorative characteristics are conferred on those less able to regulate their activity level after entering elementary school—they are described as distractible, aggressive, restless, hyperactive, and impulsive. A significant majority of these children develop serious learning problems, make marginal or failing grades, exhibit a wide range of externalizing behavior problems, and experience impaired social relationships.

ACTIVITY LEVEL AND GENDER

Anecdotal evidence has proliferated for years concerning gender differences in activity level, viz., that boys are more active than girls. The topic has undergone considerable empirical scrutiny by investigative teams throughout the world and was recently summarized in a meta-analytic review (Campbell & Eaton, 1999). Results of 46 studies comparing boys' and girls' activity level revealed that male infants are more active than female infants when measured by parent ratings, direct observations, and actometers. Gender differences in activity level become more noticeable during toddlerhood and early preschool, with boys being more active than girls regardless of whether they play with male or female friends (Tryon, 1991). The higher activity level is readily visible through the physical vigor and roughness in boys' play and may contribute to the preference of young children to play with same-gender children. These differences continue through senior high school, with boys engaging daily in vigorous activity for significantly longer periods of time relative to girls. Overall activity level, however, declines exponentially with increasing age and grade—males and females show a decline in total activity level of 69% and 36%, respectively, during school days from childhood through adolescence (Gavarry, Giacomoni, Bernard, Seymat, & Falgairette, 2003).

TECHNIQUES AND MEASUREMENT OF ACTIVITY LEVEL IN CHILDREN

Myriad approaches and techniques are available for measuring activity level in children. Some are developed specifically for clinical settings, others are better suited for educational settings, and some can be used across settings. The type of information and precision required by the clinician or investigator dictates the selection of measures. For example, several rating scales provide teacher, parent, or clinician judgments concerning a child's overall activity level in a particular setting. These same scales may be useful for monitoring changes in activity level associated with behavioral or pharmacological treatment. Their greatest attribute—cost effectiveness—is also their most significant shortcoming. Rating scales are subject to numerous influences, and at best reflect judgments about, rather than being actual measures of, activity level. Moreover, surprisingly few child rating scales have established convergent validity with objective measures such as actigraphs. Conversely, actigraphs provide highly accurate information concerning activity level and can document even minor movement differences between different body parts 24 hours a day for several weeks. Software scoring programs facilitate data analysis and interpretation. The actigraph's greatest attribute—precision of measurement—is also its most significant shortcoming. Moment-to-moment changes in activity level provide no context for understanding the functional nature of children's behavior, and most commercially available systems are prohibitively expensive. Child activity level measures and techniques reviewed next highlight these issues, and an evidence-based approach for practice is recommended.

SUBJECTIVE MEASURES OF ACTIVITY LEVEL: PARENT, TEACHER, AND CLINICIAN RATING SCALES

Practice

Rating scales are the most commonly used measures of children's activity level. Some are subscales of broadband rating instruments (e.g., the Nervous-Overactive subscale of the Teacher Report Form), whereas others are stand-alone measures of children's activity level (e.g., Werry–Weiss–Peters Activity Rating Scale) or contain the DSM-IV (American Psychiatric Association, 1994) hyperactivity items in rating scale format. Commonly used scales for measuring children's activity level are shown in Table 6.1, coupled with their psychometric properties and other characteristics.

Clinicians use activity rating scales primarily for diagnosis and treatment monitoring. The considerable weight assigned rating scale scores for determining a child's diagnostic standing is associated with several factors: (a) the unrealistic time parameters permitted by health maintenance organizations (HMOs) for assessment purposes; (b) a need to obtain information from multiple informants to assess the breadth, severity, and situational or pervasive nature of problems relative to established norms; (c) the limited insight of children concerning their behavioral and emotional problems; (d) their cost effectiveness; and (e) recognition that a child's activity level in an assessment setting fails to augur their activity level in naturalistic settings. A past study illustrates this latter point—approximately 80% percent of children meeting ADHD diagnostic criteria were misdiagnosed by their primary pediatrician because they failed to exhibit a higher-than-normal level of motor activity during the office examination. At 3-year follow-up, these children were no different from obviously hyperactive children with respect to their continuing behavior problems, poor grades, and medication status (Sleator & Ullmann, 1981).

Using activity rating scales for diagnostic purposes requires age and gender norms. Published normative data are available for some rating scales. Others, however, fail to include sex and age norms, which limit their usefulness (see Table 6.1). Most scales can be completed quickly and reflect observer judgments over time periods ranging from days to months. Test–retest reliability varies considerably and reflects scale integrity, situational variability, and the assessment time interval. Most scales demonstrate adequate short-term test–retest reliability, although there is substantial variability between the subscales of some forms. For example, the 2-week CBCL-TRF retest reliabilities range from .60 for the Withdrawn/Depressed subscale to .96 for the Total Problems measure and .93 for the Hyperactivity-Impulsivity scale. High test–retest reliability increases the likelihood that differences found during frequent administration (e.g., when monitoring changes during treatment) are due to clinical factors and not measurement error.

Internal consistency, a measure of the extent to which items on a scale measure the same underlying construct, range from an alpha of .58 to .99 for these scales. The former value may be too low, whereas the latter is also problematic, because it suggests redundancy within the item pool. Support for the validity of most available scales comes from comparisons with existing scales or demonstrations of the scale's ability to differentiate between clinical and nonclinical samples (e.g., significant group differences between ADHD and non-ADHD children). The only scales correlated significantly with objective measures are the WWPARS (with actometer ratings) and the CRS-R (with the computerized Continuous Performance Test of vigilance/sustained attention).

Age and gender norms are available for several of the rating scales. Some scales, such as the CRS-R series, provide norms for multiple age groups within their reported age range (e.g., the CRS-R provides separate norms for each gender at ages 3–5, 6–8, 9–11, 12–14, and 15–17). Others dichotomize children into larger age groupings (e.g., ages 6–11 and 12–18 for the CBCL) or provide only overall norms (e.g., WWPARS). Due to rapid changes in children's activity level, a scale that provides incremental norms is recommended for children between the ages of 5 and 10. Although some scales report the time interval rated (e.g., the CBCL asks informants to consider the child's behavior during the last 6 months), many fail to specify a rating period, an important limitation when interpreting available norms. The cost of rating scales is as variable as their psychometric properties, with starter kits ranging from $98 (CSI-4) to $425 (CRS-R) for broadband scales, and $42 (ADHD-IV) to $206 (ADDES-2) for narrow-band DSM-IV scales. The cost of reordering forms ranges from free (ADHD-IV) to more than $2 per form (RBPC). As revealed in Table 6.1, psychometric properties are not necessarily correlated with cost. Most scales can be administered quickly (range: 2–20 minutes), and some have software available for quick scoring.

Activity rating scale scores complement conventional diagnostic practice. They provide an important but limited piece of information that is juxtaposed with extensive history taking, record review, parent and child semistructured clinical interview, psychoeducational assessment results, and broad- and narrow-band scale scores. This information is coupled with details concerning the onset, course, and duration of presenting problems and serves as the basis for case conceptualization.

TABLE 6.1 Broadband, Stand-alone, and DSM-IV-Oriented Rating Scales Used to Measure Activity Level in Children

Scale		Psychometric Properties											
								Validity Evidence		Norms			
Scale		No. of Items	Item Type	Completion Time (min)	Test–Retest	Internal Consistency (alpha)	Convergent/ Divergent	Criterion	ROM	Age	Sex	Rating Period	Publisher/Cost
BROADBAND SCALES													
BASC Reynolds & Kamphaus, 1992 Hyperactivity Scale Note: BASC-2 available soon	P T	126–138	0–3 freq.	10–20	.85–.95	mid. –70s .80s–.90s	CBCL (.71–.84) CPRS Hyperactivity scale = .56	BGD: With and without CD, behavior disorder, Depression, emotional disturbance, ADHD, LD, Mild MR, and Autism	*NR*	4–18	Y	*NR*	American Guidance Services, Inc. Complete kit: $330/$450 (with or without software; manual, 25 of each form type, 25 DO forms) Forms: 25 @ $39
CBCL Achenbach & Rescorla, 2001 Nervous-Overactive scale Rater: Parent		118	0–2 freq.	15–20	(1 wk) .80–.94 (3 mo) M = .84	.63–.97	93 items identical between CBCL & TRF See manual for detailed validity information	Discriminant analysis (referred v. nonreferred: 80– 88% correctly identified BGD: ADHD v. non-ADHD, LD	NS correlation with actometer (Aronen et al., 2002)	2–18	Y	Past 6 mo.	ASEBA Complete computer scoring kit: $325 (manual, software, 50 CBCL, TRF, & YSR forms) Forms: 50 @ $25 (any type)
CBCL-TRF Achenbach & Rescorla, 2001		118	0–2 freq.	15–20	(2 wk) .60–.96	.72–.95	Factor scores "correlate well" with equiv. CTRS scales	Discriminant analysis (referred v. nonreferred	*NR*	5–18	Y	Past 2 mo.	see CBCL

Nervous-Overactive scale Rater: Teacher						(2 mo) .63–.88		Significant correlation between scores and DO	74–80% correctly identified BGD: ADHD v. non-ADHD, LD					
CBRF-A Van Egeren, Frank, & Paul, 1999 Overactivity (OA) Scale Inpatient only			65	0–3 sev.	*NR*	(2 wk) .63–.76 (3 wk) .38 (OA)	*NR*	Overactivity scale with: CBCL: .23–.28 (Externalizing, Aggressive Behavior) –.21 (Internalizing)	Overactivity scale did not predict length of hospital stay	*NR*	3–17	N	8-hour shift	Author Cost: *NR*
CBRSC Neeper, Lahey, & Frick, 1990 Motor Hyperactivity scale (4 items) Rater: Teacher			70	0–5 freq.	10–15	Published norms based on earlier 81-item of test—*NR* for available scale					6–14	Y	*NR*	Harcourt Assessment, Inc. Complete kit: $133 (contents unspecified) Forms: $29 (number unspecified)
CRS-R Conners, 1997	Long	P	80	0–3 freq.	15–20	(6–8 wk) . .47–.85	.73–.94	Long: Short: .95–1.00 P:T: .12–.55 CDI:	BGD: ADHD v. non-ADHD, emotional problems	CPT: $r =$ .33–.44	3–17	Y	Last month	Multi-Health systems, Inc. Complete kit: $425 (manual, 25 of each form type, etc.)

(continues)

TABLE 6.1 *(continued)*

Scale		No. of Items	Item Type	Completion Time (min)	Test–Retest	Internal Consistency (alpha)	Convergent/ Divergent	Validity Evidence: Criterion	Validity Evidence: ROM	Norms: Age	Norms: Sex	Rating Period	Publisher/Cost
							Psychometric Properties						
Hyperactivity Scale	T	59			.47–.88	.77–.96							
Rater: Parent, Teacher	P	28					.40–.82 (teacher)						Forms: 25 @ $26 (any version)
Self-report and							.36–.79 (parent)						
ADHD DSM-IV	Short			5–10	.89–.96	.86–.94							
Scales also available	T	27		.72–.92	.88–.95								
CSI-4	P	77	0–3 freq.	10–20	(2 wk)	.74–.94	ADHD-HI & (C) vs:	ADHD category (6 studies):		5–12	Y	"Overal 1 bx"	Checkmate Plus, Inc.
Gadow & Sprafkin, 1994, 2002	T	99			.66–.88 (P & T)	(both)	TRF-Externalizing: .69 (.53)	Mean sensitivity: .64 (.87 if using either teacher or parent)					Complete kit:
Screens for DSM disorders, including ADHD							IOWA-Conners-IO: .70 (.83)	Mean specificity: .77 (.62 if using either teacher or parent)					$98/$358 (with or without software; manuals, 25 each form, plus scoring sheets, profiles) Forms: 50 @ $32
MCBC	P	77	T-F	10–15	(unknown)	.58–.83	*NR*	*NR*	*NR*	9–14	Y	*NR*	Author
Sines, 1986, 1988					.49–.78								Cost: *NR*
Activity Level Scale	T	68			*NR*	*NR*							

PBQ Behar & Stringfield, 1974 Hyperactive-distractible factor		30	0–2 freq.	5–10	(3–4 mo) .60–.94	*NR*	Significantly correlated with DO of classroom bx and interactions	BGD: Normal v. hyperactive, emotionally disturbed preschool children	*NR*	3–6	*NR*	*NR*	Author Cost: *NR*
RBPC Quay & Peterson, 1993 Motor Excess (ME) Scale (5 items)		89	0–3 sev.	15–20	(2 mo) .49–.83	.73–.94	Original BPS: .63–.97 CBCL: .43–.92 among alike scales ME not significantly correlated with DO of gross motor activity ($r = .00$; $N = 34$)	Discriminate clinical from nonclinical groups of children	*NR*	5–18	N	*NR*	Psychological Assessment Resources, Inc. Complete kit: $172 (manual, 50 forms & profile sheets) Forms: 25 @ $55
STAND-ALONE SCALES													
WWPARS Werry, 1968 Rater: Parent		22	0–2 freq.	5	*NR*	*NR*	WWPARS rating at age 4.5 significantly related to CPRS rating at 6.5 years (Campbell et al., 1978)	Predicted some improvements in mother–child interactions in response to stimulant meds in hyperactive children (Barkley & Cunningham, 1980)	Actometer M = .65 Range: .24 (classroom) to .77 (woodshop); .67 during gym	2–9	*NR*	*NR*	Free online:
DSM-IV SCALES													
ACTeRS	P	25	1–5 freq.	5–15	(2 wk)	.78–.96	CBCL (.80–.81)	Hyperactivity scores medication sensitive in ADHD students; BGD: ADHD/non-ADHD	*NR*	5–14	N	*NR*	MetriTech, Inc. Complete kit: $54 (manual + 50 forms; parent OR teacher)

(*continues*)

TABLE 6.1 *(continued)*

Scale		No. of Items	Item Type	Completion Time (min)	Test–Retest	Internal Consistency (alpha)	Convergent/ Divergent	Validity Evidence: Criterion	Validity Evidence: ROM	Norms: Age	Norms: Sex	Rating Period	Publisher/Cost
Ullmann et al., 1997 Hyperactivity scale	T	24			.68–.78	.92–.97	CPRS (.78–.90)						Forms: 50 @ $37
ADDES-2	P	46	0–4 freq.	15–20	(1 mo)	.96–.98	TRF (.79–.87)	BGD: ADHD v. non-ADHD	*NR*	3–20	Y	*NR*	Hawthorne Educational Services, Inc.
McCarney, 1995a, 1995b					.88–.93								
Hyperactive/impulsive subscale Note: ADDES-3 available soon	T	60			.88–.97	.98–.99	CTRS: –2: (.51–.83) -S: (.42–.80) -S: ADHDT (.76–.93) -S: TRF (.51–.87)			4–19			Complete kit: $206 Forms: 50 @ $31 (either version) $20 per 50 ADDES/DSM-IV forms
ADHD-SC-4	P	50	0–4 freq.	*NR*	(6 wk)	.93–.95	CBCL (.48–.81)	Sensitivity	*NR*	3–18	Y	*NR*	Slosson Educational Publications, Inc.
Gadow & Sprafkin, 1997a, 1997b					.75–.82		TRF (.45–.88)	.81–.85 Specificity .57–.94					Complete kit: $78 Forms: 50 @ $28
	T				.70–.89	.92–.95		Sensitivity .61–.89 Specificity .57–.94					

ADHD-IV DuPaul, Power, Anastopoulos, & Reid, 1998	P	18	0–3 freq.	10	(4 wk) .78–.86	.86–.92	CPRS-R: .10–.80 CRTS-R: .12–.41 ADHD Bx Code: .25–.26	65–84% (teacher); 60–68% (parent) of ADHD children correctly classified via logistic regression		5–18	Y	Last 6 mo.	Guilford Publications, Inc. Complete kit: $42 (book, manual, scale)
Hyperactivity/ Impulsivity scale	T				.88–.90	.88–.96	AES: −.36	BGD: ADHD-I, ADHD-C, and controls (parent and teacher)					Forms: Photocopiable
ADHD-SRS Holland, Gimpel, & Merrell, 1998 Hyperactive-Impulsive Scale Rater: Parent, teacher (same form; separate norms provided)		56	0–4 freq.	10–15	(2 wk) .93–.98	.92–.99	ADDES: .80–.95 CTRS: .64–.97 (like scales) ADHD-IV: .89–.93	BGD: ADHD v. non-ADHD	*NR*	5–18	Y	*NR*	Wide Range, Inc. Complete kit: $85 (manual, 25 forms) Forms: 25 @ $38
ADHDT Gilliam, 1995 Hyperactivity scale Rater: Parent or Teacher		36	0–2 sev.	5–10	(1 wk) .85–.94 (2 wk) .85–.92	.91–.97	Compared to seven other tests used to diagnose ADHD or bx concerns: "satisfactory" CTRS: .53–.72 ADDES-S: −.81 to −.88 ACTeRS: −.71 to −.78	BGD: ADHD v. non-ADHD True positives/ negatives: 92% False positives: 7.7%	*NR*	3–23	Y	*NR*	Pro-Ed, Inc. Complete kit: $95 (manual, 50 summary/response forms) Forms: 50 @ $42

(*continues*)

TABLE 6.1 *(continued)*

Scale		No. of Items	Item Type	Completion Time (min)	Test–Retest	Internal Consistency (alpha)	Convergent/ Divergent	Validity Evidence: Criterion	Validity Evidence: ROM	Norms: Age	Norms: Sex	Rating Period	Publisher/Cost
					Psychometric Properties								
BASC-M	P	47	0–3 freq.	5–10	(2–8 wk)	.64–.83	CBCL: –.68 to .79	BGD: ADHD v. non-ADHD &	*NR*	4–18	Y	*NR*	American Guidance Services, Inc.
Reynolds, & Kamphaus 1992					.57–.90			ADHD-I v. ADHD-C (both versions)					Complete kit: $129 (manual, 25 each form, scoring template)
Hyperactivity Scale	T				.72–.93	.78–.93	CTRS-R (–.36–.62)						Forms: 25 @ $32
CAAS	P	31	1–4 freq.	2–5	(3 yr)	.75–.81	*NR*	*NR*	*NR*	5–13	*NR*	*NR*	American Guidance Services, Inc.
Lambert, Hartsough, & Sandoval, 1990					.32–.44								Complete kit: $150 (manual, 25 each form, scoring profile)
Hyperactivity scale	T				.40–.82	.78–.94							
Rater: Teacher or parent													Forms: 25 @ $48
ECADDES	P	50	0–4 freq.	12–15	(1 mo)	.87–.97	CPRS-R: .41–.71	BGD: ADHD v. non-ADHD	*NR*	2–6	Y	*NR*	Hawthorne Educational Services, Inc.
McCarney & Johnson, 1995					.82–.89		ADHDT: .46–.81						Complete kit: $162 (technical and intervention manuals, parent's guide, 50 school, home, & DSM-IV forms)
Hyperactive-Impulsive Index	T	56			.90–.98								Forms: $35 per 50 forms (home or school); $22 per 50 DSM-IV

S-ADHD-RS Spadafore & Spadafore, 1997 Impulsivity/ Hyperactivity scale (20 items) Rater: Teacher	50	0–4 freq.	10	(2 wk to 1 mo) .88–.90	*NR*	*NR*	Correctly identified: 50/50 ADHD Incorrectly identified 2/50 non-ADHD as ADHD	*NR*	5–19	Y		Complete Kit: $80 (manual, 25 Scoring Protocols, 25 Observation Forms, 25 Med Tracking Forms) Forms: 25 @ $25
SNAP-IV Swanson, 1992 Hyperactivity/ Impulsivity scale (9 items) Rater: Teacher or parent (same form)	90	0–3 freq.	10–15	(2 wk) .70–.90	.90+	*NR*	*NR*	*NR*	*NR*	Y	*NR*	Form and scoring guidelines are free online: http://www.adhd.net

Abbreviations: ACTeRS, ADD-H: Comprehensive Teacher's Rating Scale; ADDES-2, Attention Deficit Disorders Evaluation Scale—2nd edition; ADDES-S, Attention Deficit Disorders Evaluation Scale—Secondary-Age Student; ADHD-IV, ADHD Rating Scale-IV; ADHD-SC-4, ADHD Symptom Checklist—DSM-IV edition; ADHD-SRS, ADHD Symptoms Rating Scale; ADHDT, ADHD Test; BASC, Behavior Assessment System for Children; BASC-M, Behavior Assessment System for Children—Monitor for ADHD; BGD, Between-group differences; CADS, Conners' ADHD/DSM-IV Scales; CAAS, Children's Attention and Adjustment Survey; CBCL, Child Behavior Checklist; CBRF-A, Child Behavior Rating Form—Abbreviated; CBRSC, Comprehensive Behavior Rating Scale for Children; CDI, Children's Depression Inventory; Con., Convergent; CPRS-R, Conners' Parent Rating Scale—Revised; CPT, Continuous Performance Test; CRS-R, Conners' Rating Scales—Revised; CSI-4, Child Symptom Inventory—4th edition; CTRS, Conners' Teacher Rating Scale—Revised; ECADDES, Early Childhood Attention Deficit Disorders Evaluation Scale; freq., frequency; M, mean; MCBC, Missouri Children's Behavior Checklist; *NR*, P, Parent; PBQ, Preschool Behavior Questionnaire; RBPC, Revised Behavior Problem Checklist; ROM, relation to objective measures; S-ADHD-RS, Spadafore ADHD Rating Scale; sev., severity; SNAP-IV, Swanson, Nolan, & Pelham Rating Scale—Version IV; T, Teacher; TRF, Teacher Report Form; WWPARS, Werry–Weiss–Peters Activity Rating Scale.

Pricing information obtained from publishers' Web sites, retrieved January 2005.

Another common use of children's activity rating scales is for monitoring treatment effectiveness. Teachers and parents frequently complete short-form activity scales during the initial titration period to inform physicians of pharmacological treatment effectiveness. Instruments such as the Conners' Teacher Rating Scale (CTRS; Conners, Sitarenios, Parker, & Epstein, 1998), SNAP-IV (Swanson, 1992), ACTeRS (Ullman et al., 1997), and ADHD-IV (DuPaul, Powers, Anastopoulos, & Reid, 1998) are useful for this purpose because of their proven sensitivity to overall and between-dose psychostimulant effects (Rapport, 1990). These and other scales are also useful for monitoring changes in activity level that accompany behavioral interventions (e.g., Rapport, Murphy, & Bailey, 1982; Multimodal Treatment Study of Children with ADHD Cooperative Group, 1999).

Research

Activity rating scales are commonly used in research studies for a variety of purposes. Some studies use a particular score (e.g., ≥2 SD above the mean) as one of several selection criteria for identifying sample participants. Scales with published norms (age, gender) and reasonably high sensitivity and specificity are best suited for this purpose (see Table 6.1).

Other researchers obtain rating scale scores to investigate relationships between children's activity level and other characteristics, such as aggression, as potential marker variables of particular outcomes, or as mediators and moderators of other variables under study. Examples of these studies are common in the developmental psychopathology literature, whose central question addresses frequently the identification of early child characteristics and their continuity with later adverse outcomes. Fergusson, Lnyskey, and Horword (1997), for example, found that early ratings of hyperactivity predict adverse scholastic achievement during adolescence, even after controlling for IQ and other family background factors. These findings were replicated by an independent investigative team and were expanded to show that a dual-developmental model involving cognitive as well as behavioral pathways accounted more fully for the relationship between early attention deficit and long-term scholastic achievement (Rapport, Scanlan, & Denney, 1999). Scales that can be readministered over extended time intervals and have strong test–retest reliability and internal consistency are preferred in these types of studies.

Gauging treatment success by obtaining measures before and after behavioral or psychopharmacological treatment is a third common use of activity rating scales in child research. Rating scales used for assessing treatment outcome need to have high test–retest reliability over brief periods of time and with repeated administration (>.70), demonstrated sensitivity to treatment-related activity level change, and scale construction that minimizes floor and ceiling effects. Several scales possess these characteristics (see Table 6.1), and some are exquisitely sensitive in detecting overall and between-dose effects (see Rapport, 1990).

Limitations

Inadequate measurement units, insufficient psychometric properties, and the lack of age and sex norms limit the usefulness of many child activity rating scales. None of the available child activity level rating scales has a standard unit of measurement—that is, a meaningful unit of activity or movement that is equivalent within and across measures and raters. This lack of a basic measurement unit renders it impossible to accurately define activity level and leaves us in the uncomfortable position of referring to differences and changes in children's activity level by referring back to the scales used initially to quantify the behavior. The fact that many available scales are moderately or even strongly correlated with one another partially mitigates this concern. Correlations between activity rating scales and objective activity measures, however, are generally much weaker and in the .32–.58 range. These values indicate that 66% to 91% of the variability in activity rating scale scores is not linearly related to variability in actigraph scores in the same children measured at the same time. This finding probably reflects the fact that children's activity rating scales tend to reflect other aspects of behavior and not just activity level—an expected circumstance, given the wording of most scale items and reliance on factor analytic scale construction methodology (i.e., descriptions of activity level may correlate highly but tend to reflect a broader range of behavior than just movement).

Psychometric limitations of rating scales include imprecise measurement, inadequate construct validity, and a lack of established research documenting their diagnostic utility. Most scales use a Likert-type rating format, and comparisons within and between scales assume that a specific interval reflects the same unit of behavior (i.e., that the unit of measure between a 2 and 4 is identical to the difference between a 1 and 3 and that these behavioral units are consistent across scales). This is a speculative assumption at best. Moreover, the psychometric properties of most child activity rating scales are wanting. Activity scale validity is usually accomplished by demonstrating that scores derived from one instrument correlate with those derived from an already established scale of activity level or by demonstrating that children known to be highly active (e.g., ADHD) score significantly higher than normal peers on the scale. Most activity rating scales meet at least one of these two criteria. Extant research, however, reveals that even when activity scale scores are correlated, they may be unrelated to objective measures of activity level (Rapoport, Abramson, Alexander, & Lott, 1971; Stevens, Kupst, Suran, & Schulman, 1978). Other studies demonstrate that children receiving higher teacher activity ratings than other children in the same classroom may actually be less motorically active, according to precision counters used concurrently to measure motor movement. For example, when measured by step counter, nearly 64% of children rated as clinically hyperactive were less active than the most active child rated as being normal by the teacher (Tryon & Pinto, 1994). Collectively, these shortcomings limit the interpretability and usefulness of activity rating scales.

The diagnostic utility of most activity rating scales is unknown. Four metrics address this concern: sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). Sensitivity and specificity indicate the proportion of the group with a target diagnosis who test positive and negative on a measure, respectively. These two indices are useful for examining the overall classification accuracy of rating scales and other instruments but are not particularly valuable to clinicians unaware of a child's diagnostic standing prior to referral. PPP and NPP are the statistics most relevant for this purpose. PPP, as it applies to rating scale utility, indicates the conditional probability that a child exceeding a rating scale cutoff score meets criteria for a particular diagnosis such as ADHD (i.e., the ratio of true positive cases to all test positives). NPP, in contrast, indicates the conditional probability that a child who doesn't exceed an established cutoff score will not meet criteria for a particular clinical diagnosis (i.e., the ratio of true negative cases to all test negatives). High values (e.g., >.80) for all four indices are desirable.

Sensitivity, specificity, PPP, and NPP are also used to examine individual scale items and specific clinical diagnostic criteria. For example, the diagnostic utility of DSM-III (1980) activity level descriptions was investigated in a study of 76 6- to 12-year-old boys referred to a child psychiatry outpatient clinic (Milich, Widiger, & Landau, 1987). Children met diagnostic criteria for ADHD, Conduct Disorder (CD), both diagnoses, some other diagnosis, or an unspecified deferred diagnosis based on a comprehensive clinical assessment. Base rates for DSM-III items "shifts activities," "runs/climbs," "can't sit still," "runs around," and "on the go" ranged from .25 to .51 in the children with ADHD. As such, these descriptions occurred with relatively low frequency, and only two (shifts activities, on the go) identified a reasonably large proportion of the children (i.e., sensitivity rates of .70, .40, .38, .38, .68, respectively). The items were fairly specific to the disorder (specificity rates of .69, .78, .89, .86, .67, respectively) and served as moderately strong inclusion criteria (PPP rates of .72, .67, .79, .75, .69, respectively). NPP rates ranged from .54 to .68 and indicate that none of the items are particularly useful as exclusion criteria—that is, the absence of the symptoms do not necessarily mean that ADHD can be ruled out. Collectively, these findings support current clinical practice parameters that recommend conducting a comprehensive clinical assessment as opposed to relying on activity rating cutoff scores.

A final limitation of many child activity rating scales is their failure to include information concerning temporal changes in activity level relative to age, gender, and cultural norms. This is important because children exhibit differences or changes in activity level for myriad reasons, and some patterns are associated with particular psychological disorders, diseases, and situational demands. Complementary information is always needed, such as whether the change is abrupt or gradual, situational or pervasive, acute or chronic, and stable or waxing and waning.

Recommendation

Activity rating scales play a vital role in clinical diagnosis and treatment monitoring. Astute clinicians and researchers recognize that scales with established psychometric properties provide important information concerning children's activity level if they include age and gender norms and are culturally appropriate. Reviewing a scale's psychometric properties is particularly important when assessing low rates of activity level such as those observed in some cases of childhood depression—reliability (and hence, validity) tends to be lower in these cases. Monitoring treatment effects with activity rating scales requires prudence. Diminished activity level in children may or may not correspond with positive change in other aspects of their behavior. Scales such as the Academic Performance Rating Scale (APRS: DuPaul, Rapport, & Periello, 1991) are recommended to obtain complementary information concerning corresponding changes in adaptive functioning.

OBJECTIVE MEASURES OF ACTIVITY LEVEL: ANALOGUE, ACTOMETER, ACTIGRAPH, AND BEHAVIOR OBSERVATION SYSTEMS

Practice

Objective measures of children's activity level have been available for decades, but few make their way into everyday clinical practice. The primary reasons for their infrequent adoption are cost, time demands, and the widely held belief that they provide diminutive incremental benefit in diagnostic accuracy relative to information derived from parent and teacher activity rating scales. Several of the newer measures are less expensive and require minimal time to download collected data, and recent empirical evidence suggests that some instruments and observational systems hold excellent promise for improving diagnostic sensitivity and specificity. Objective measures of children's activity level include analogue measures, behavioral observation systems, actometers and actigraphs, and other measures used exclusively in research settings, such as stabilometers, photoelectric cells, ultrasound, and infrared motion analysis (see Table 6.2).

Analogue measures of children's activity level are occasionally used in research-oriented clinical outpatient facilities. An area within the clinic is established that approximates some aspect of children's natural environment, with the expectation that evoked activity level will resemble behavior observed typically in *esse situ*. This setting can take many different forms, depending on the information desired. Clinicians measuring children's movement associated with avoidance of or escape from a feared stimulus, such as a spider, will frequently utilize an elongated *approach* corridor and record steps taken or grid lines crossed as an indicator of fearfulness. Other common applications involve the use of analogue classrooms or activity rooms to assess motor movement in children referred for

ADHD, and topographical movement idiosyncrasies in children with pervasive developmental disorder. Activity level in these settings is usually assessed by trained observers *in vivo* or afterwards by tape review. Measurement may be as simple as counting floor marked grid changes over a preestablished time interval (e.g., Partington, Lang, & Campbell, 1971) or as sophisticated as using a computer-based, behavioral observation system to review taped sessions and code multiple behavior categories.

Direct observation has several advantages over other methods for measuring children's activity level. Some observational coding schemas are highly correlated with mechanical measures such as actometers, motion sensors, and calorimetry (Tryon, 1991). They do not hinder a child's movement as do some mechanical devices. In addition, they can be used to reflect specific types of movement in a variety of settings and over defined time periods. This latter information is particularly important if the goal is to understand the functional nature of children's movement rather than simply measuring quantitative differences. Most observation coding schemas, however, require that one or more trained individuals observe a child for a designated time interval (e.g., 20–30 minutes) on several occasions within or across weeks to render a valid sampling of behavior. For this reason, they are impractical for most clinicians.

An interesting alternative to *in vivo* observations is to digitally record the child's behavior in the setting and situation of interest and to transfer this information to a computer mpeg file for later viewing and analysis via the Noldus Observer (Noldus Information Technology, Inc.). This sophisticated software program permits continuous recording of any behavior while time-locking the observation with noted changes in the environment. Additional features include the ability to view, stop, and reverse the observation at any time at significantly reduced speeds, to enhance accuracy, and to modify behavioral codes for additional analyses. Newer versions and complementary software facilitate the exploration of sequential behavioral events that may not be immediately apparent to a trained observer—for example, to determine whether particular behaviors systematically precede other behaviors. A screen shot of the Noldus Observer is depicted in Figure 6.1.

Pedometers are inexpensive instruments used occasionally to measure one aspect of children's activity level—steps taken. Past studies, however, indicate that actometers are much more sensitive than pedometers for both detecting overall activity level and discriminating between children with and without ADHD (Barkley & Ullman, 1975; Rapoport et al., 1971).

Actometers are modified wristwatches capable of recording the frequency of movement occurring along the horizontal plane crossing through 3 and 9 o'clock on the watch (Bell, 1968). Removing the balance and hairspring assemblies and adding a small weight to the lever causes the second hand to move incrementally each time the device is moved. Frequency of movement, or how many distinct movements occur (i.e., changes in direction), is estimated by comparing the starting and ending times shown on the watch face. These instruments, however, are

TABLE 6.2 Objective Measures of Activity Level in Children: Mechanical and Direct Observational Approaches

Instrument/ Distributor	Age Range	Recording Length	Norms	Software Available	Cost
MECHANICAL					
Actigraphs Ambulatory Monitoring MiniMitter MTI, Inc.	Any	22 days per 32 KB of memory	No	Yes	Starter: $1000+ (with necessary software and reader interface); $500–$2000 for each additional actigraph
Actometers[a] Model 108 Engineering Department Times Industries, Waterbury, CT 06720	Any	Variable	No	Yes[b]	*NR*
Pedometers (available at sporting goods stores)	Any	Range: 99,999 steps (~5.25 miles) to 1000 miles	No		
Stand-Alone				No	$10–$40
With data downloadable to PC				Yes	$125–$400+
DIRECT OBSERVATIONS					
ADHD BCS Barkley, 1990	*NR*	15 min.	No	No	*NR*
ADHD-SOC Checkmate Plus, Ltd.	School age	16 min.			Kit: $25
BASC SOS AGS, Inc.	School age	15 min.	No	Yes	25 forms @ $33
COC Abikoff, 1977/1980	School age	32 min.	No	No	*NR*
DOF ASEBA	5–14	10 min.	Yes	Yes	*NR*
Noldus Observer Noldus Information Technology	Any	Variable	No	Yes	Observer Basic 5.0 $1795 Observer Video Pro 5.0 $5850

Abbreviations: BCS, Behavior Coding System; COC, Classroom Observation Code; DOF, Direct Observation Form; SOC, School Observation Code; SOS, Student Observation System;

[a]Many studies report either using the Kaulins and Willis actometers (no longer manufactured) or enlisting a jeweler to modify a self-winding wristwatch, as described by Schulman and Reisman (1959).

[b]Eaton, McKeen, & Saudino (1996) provide SAS syntax for performing group-level data analysis based on actometer readings.

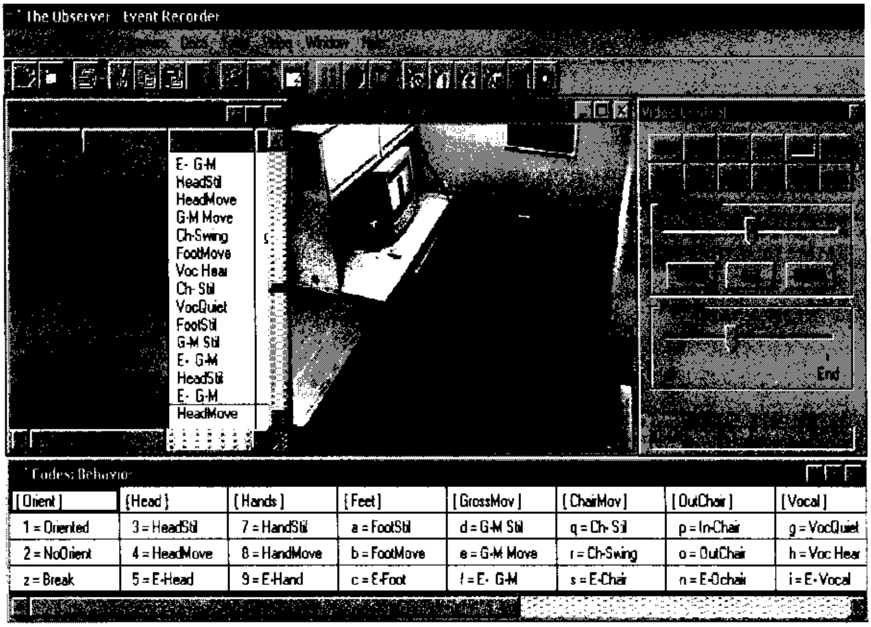

FIGURE 6.1 A screenshot of the Noldus Observer. Trained observers watch the video file (center) and code behavior (top left) according to prespecified behavior codes (bottom left).

incapable of measuring the duration or intensity of movement, which may be important for clinical diagnosis.

Newer *actigraphs* generate a current (voltage) each time the instrument is moved in a direction vertical to the instrument face. The current is passed through an amplifier and filtered based on factory and user settings. The result of this process is an analogue waveform—a histogram of measured voltage over time—from which movement frequency, intensity, and duration is extracted (for a detailed review, see Tyron, 1991). These devices allow for more precise recording of activity or movement, with excellent test–retest stability. An oft-cited actigraphy study examined movement differences in 12 children with ADHD and 12 age-matched classmates based on 24-hour waist activity for 7 days (Porrino et al., 1983). Children with ADHD were found to be approximately 25% to 30% more active than normal controls, particularly in school and at home when completing in-seat academic assignments. Deciding on where to place the actigraph (e.g., waist, arm, leg) and which recording modes to utilize are critical elements that must be considered prior to clinical assessment. These and other relevant parameters are reviewed below.

Settings and Recording Modes

Actometers do not have programmable settings, and they report only an average frequency of movement. In contrast, actigraphs provide many options for data collection, and the settings and mode of an actigraph determine the type of data recorded. Different modes return significantly different data points, and careful consideration of the purpose of data collection may prove critical when selecting among possible settings. The primary distinction is between actigraph placement and variable sampling intervals and among frequency, duration, and intensity of movement selections—these and other dimensions are reviewed next.

Placement of the actigraph influences the amount and type of movement it records. Common placement locations are the wrist, trunk, and ankles. Actigraphs placed on a belt or in a pouch around the waist are more likely to detect gross body movements, due to the motion and energy expenditure necessary to move this area. In contrast, arm recordings detect lower-intensity, more subtle movements and result in higher activity recordings, regardless of the time of day. Precise placement on the arm is also worth noting because even small variations may yield differences in activity level. Some research on the older, less precise actometer suggests that larger readings are typically obtained as a function of how far the recording site is from the axis of rotation (Johnson, 1971). Other researchers, however, have found that forearm length is not significantly related to obtained movement counts (Eaton, 1983). Thus, wrist recordings may detect higher activity than will forearm recordings, highlighting the importance of consistent placement across subjects or patients. Either wrist may be used for recording (i.e., handedness does not appear to affect differentially recorded activity levels in children). As a general rule of thumb, integrated generalized movements like postural shifts are detected by all sites, whereas small movements associated with distal extremities are best detected by wrist-worn monitors.

Filters involve measurement parameters such as sampling rate and epoch length. Sampling rate, measured in hertz (Hz), determines how many samples per second are taken by the actigraph. This rate is factory set at 10 Hz (i.e., 10 samples per second) for many actigraphs, although some models are programmable. In the most general sense, more precise data are garnered as a function of higher sampling rates (i.e., more data sampled per time unit). An *epoch* is the unit of time that the actigraph combines into one data point. Some models have factory-set epochs (e.g., Ambulatory Monitoring's MicroMini is preset at 1-minute epochs), whereas others are customizable. Epoch length represents a trade-off between specificity and duration of continuous monitoring—the shorter the epoch, the more specific the data. For example, a 1-minute epoch at a 10-Hz sampling rate over 60 contiguous minutes will produce 60 distinct data points, with each point based on 60 samples. Changing the epoch length to 10 minutes yields only six data points for the same length of observation (each based on 600 samples). Most modern actigraphs have at least 32K of memory, which translates into approximately 22 days of continuous data using 1-minute epochs. This memory is reused once data have been downloaded from the actigraph, indicating that short epochs are feasible for all but the longest continuous measurements. The nature of the data recorded for each point, however, varies considerably based on the actigraph's mode.

Mode is set depending on whether the user is interested in frequency, duration, and/or intensity of movement. Common available modes include full-waveform, zero-crossing, time above threshold, and proportional integrating measure modes. A visual heuristic of data characteristics associated with different settings is depicted in Figure 6.2. Full-waveform mode collects the entire analogue waveform for analysis, whereas the other options record only certain characteristics of this wave. This mode is typically available only on high-end actigraphs. Zero-crossing mode (ZCM) records the number of times the waveform voltage crosses a set reference voltage—as the voltage changes in response to movement, ZCM counts each instance the waveform crosses a preset threshold. Consider a swing on a child's swingset held at a determined height and released. ZCM counts the number of times the swing crosses a height just above its resting point. As the swing slows, it continues to cross this threshold, and each pass is counted until it stops. Zero-crossing mode is thus a measure of movement frequency.

Time above threshold (TAT) measures the *duration* of movement, or time spent in movement. Once the set threshold is reached, the actigraph counts each sampling period (e.g., every tenth of a second at 10 Hz) until the movement-generated voltage falls below this level. This mode determines that movement is above a threshold, but it does not measure *how much* above threshold. Consider the child's swing again—as it slows, less and less time is spent above any given height, and the number of counts will be considerably less than would be the case if the same swing were measured using ZCM. For this reason, zero-crossing is often considered a more sensitive measure than TAT.

Proportional integrating measure (PIM) mode measures the absolute value of the area under the waveform curve or, stated differently, the *intensity* of movement or activity level. To illustrate the difference between frequency and intensity of movement, consider the *test of strength* machines at state fairs. Two individuals may swing the hammer in a nearly identical manner, however, the intensity or force can vary considerably—PIM mode detects these differences. PIM mode measures gross activity level instead of counting time above threshold or merely changes in movement. Two subgroups, low PIM and high PIM, are available. The difference is that high PIM applies an extra filter, which further amplifies the analogue waveform signal and allows comparison of more minute movements. Unless one is measuring premature infants or individuals with movement-related disabilities, the standard low-PIM mode is preferred. Using high PIM with normal or hyperactive populations may result in a ceiling effect because much of the activity may be amplified to the maximum value possible for the actigraph. Hypothetical differences in data counts across the three modes are depicted in Figure 6.2.

Mode Comparisons

Zero-crossing and time-above-threshold options quantify how often movement occurs—in contrast, PIM mode examines how *much* movement occurs. PIM mode may also be the most sensitive measure, especially when examining

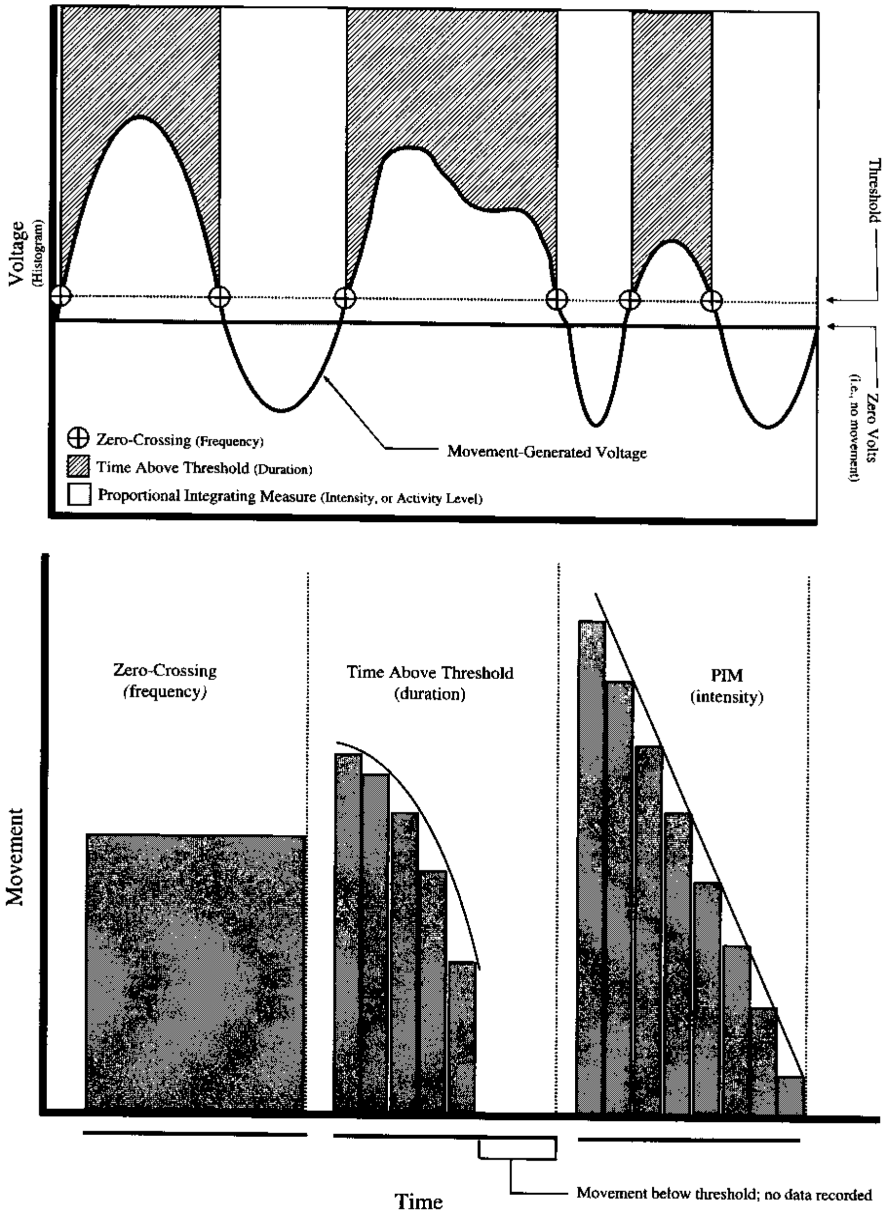

FIGURE 6.2 A visual heuristic depicting the movement-generated waveform characteristics associated with different modes (top), and hypothetical data for three actigraph modes measuring the same swing from start to stop (bottom).

differences between groups or within individual clients over time. Mathematically, the values that can be obtained for each 1-minute PIM epoch range from 0 (no movement) to 32,000 or 65,535, depending on the model of actigraph. For zero-crossing and time-above-threshold modes, the range of obtainable values for a 1-minute epoch (at 10-Hz sampling rate) drops to 0–600 (10 samples per second × 60 seconds). Realistically, the maximum for zero-crossing is likely much lower when measuring human activity, since a value of 600 would correspond to changing direction of movement every tenth of a second. The choice of one or more of these options depends on the referral or research question. If one is interested in activity level, PIM mode may be preferable for its ability to quantify the gross amount of movement. With some newer models, two or even all three of these modes can be recorded simultaneously (at a cost to length of observation), allowing the clinician or researcher to examine multiple aspects of movement, thus providing a richer basis for diagnosis or comparison. PIM mode may be beneficial when examining children with ADHD. Conversely, a ZCM frequency count may be preferable for examining medication effects on stereotypic behavior in autism. In the latter case, the intensity and duration of each movement may change minimally, and a frequency count may reveal more precisely whether the movements are occurring less or more often as a function of treatment.

Research

Objective measures of activity level are used in research for myriad purposes. Pedometers, actometers, and actigraphs are commonly used to study activity changes in infants and normally developing children, in sleep studies as a nonspecific sign of neurological conditions such as Alzheimer's and Parkinson's, and as outcome measures in weight loss and exercise studies (for a review, Tryon, 1991). The estimated reliability for actigraphs placed at the same site in the same person range from .90 to .99 (Tryon, 1985), and actigraph data are highly correlated with children's playroom activity level recordings (Ullman, Barkley, & Brown, 1978). Behavioral observations also tend to be highly correlated with actometers; however, caution must be exercised in certain situations—lower estimates are obtained during specific types of activity (Halverson & Waldrop, 1973) and in low activity level recordings in general.

Every conceivable type of objective measure has been used to study children with ADHD—inexpensive pedometers and actometers to technologically sophisticated measures such as actigraphs, photoelectric cells, ultrasound, and infrared motion analysis. Measurement selection depends on the research question and on the level of precision required by the investigator. Studies of ADHD illustrate this point adeptly. Simple actometers placed on ankles and wrists and in waist belt pockets consistently reveal overall activity level differences between children with ADHD and control children. More sophisticated actigraphs are required to determine whether these children move more than their peers throughout the day and during different activities (e.g., Porrino et al., 1983). These devices, however, are unable to address questions concerning the type, site, and complexity of movement. For example, Teicher et al. (1996) used infrared motion analysis to track the precise two-dimensional location of four reflective markers taped to a cap and clothing worn by children with ADHD and normal controls. Children with ADHD exhibited two to three times more movement during a vigilance task at every body location, and covered a fourfold-wider area in their movements relative to control children. Motion analysis also revealed that children with ADHD moved their entire body to a greater degree, kept their extremities still 66% less and their trunk still 74% less, and exhibited less complex movement overall (i.e., more linear side-to-side movement) relative to control children. Information gleaned from this investigation confirms that children with ADHD emit significantly more movement relative to age-matched controls, regardless of body location monitored. It also reveals that the excessive movement is less complex relative to same-age peers. But why do children with ADHD display a higher rate of movement? And are these excesses merely random movement, or do they serve some functional purpose? Complementary studies suggest that their excessive movement may be related to particular task demands and, in some cases, serve to compensate for an inability to maintain sufficient arousal (Zentall & Leib, 1985). These and related questions must be addressed by behavioral observation and complex coding schemas.

Research has also sought to determine whether actigraphs are diagnostically useful given their exceptional reliability and established validity. Studies have not been designed to address definitively whether they can serve this function, but some have examined pieces of the puzzle. For example, a belt-worn activity monitor showed exceptionally high positive predictive power (.91) for discriminating between children with ADHD and normal controls undergoing psychometric evaluation (Matier-Sharma, Perachio, Newcorn, Sharma, & Halperin, 1995). Sensitivity, however, was unacceptably low (.25). Collectively these findings indicate a very strong likelihood that children will meet ADHD diagnostic criteria if they exhibit higher-than-normal activity during psychometric evaluation, whereas nearly 75% of children with a diagnosis of ADHD fail to show this behavioral pattern. Conclusions derived from this and similar studies are complicated by methodological factors. The aforementioned cited study, as an example, relied on the least sensitive placement for detecting activity level changes—the waist, which primarily detects major postural shifts.

Limitations

Even objective measures of children's activity level have relative strengths and shortcomings. It is unclear whether analogue test settings elicit a representative sample of children's behavior shown at school and home. This limitation raises questions about the validity and generalization of findings. The relatively brief behavioral samples of 10–30 minutes in analogue settings are particularly suspect. Children most likely to be assessed for excessive movement in analogue settings are those suspected of ADHD. These children are known to be highly variable, which decreases the likelihood that brief behavior samples will serve as adequate behavioral samples for either diagnostic or treatment purposes.

Direct observation bypasses many of the shortcomings inherent in analogue assessment but tends to be quite costly, even when volunteers are recruited and trained. Required training time is proportional to the complexity of the coding scheme and behavior being observed. Periodic retraining is required to minimize *observer drift*—a well-documented phenomenon that refers to the likelihood that observers will alter their definitions of the behaviors they are observing over even short time periods. Taping observations can dramatically improve reliability by creating a permanent record of behavior, but observers must still be trained to review and code the data reliably. Sophisticated observational and data-handling programs represent a clear improvement over existing observation instruments and procedures, but they remain too costly for most practice settings. High-tech instruments are procured to conduct investigative studies of children's activity level and usually represent a trade-off between measurement precision and transportability. The stabilometer lies on the simpler end of the spectrum and provides highly defined measurement of a child's buttocks movement while seated. The very nature of the apparatus, however, places serious limits on a child's mobility and may offer a limited behavioral sample. Pedometers are limited to providing somewhat crude measures of steps taken, whereas actigraphs allow for more precise recording of activity or movement over extended time periods, with excellent test–retest stability. Their most significant limitations are that they are limited to detecting movement of particular extremities, due to the attachment locale or site; they remain too costly for widespread clinical adoption; and age, gender, and cultural norms are unavailable. Instruments at the upper end of the measurement spectrum, such as ultrasonic, are somewhat unreliable in detecting gross and fine motor movements—others, such as infrared motion analysis, are prohibitively expensive and must be supplemented by observation to judge the functional nature and contextual variables associated with movement.

Recommendations

Behavioral observation is an expensive and time-consuming practice unlikely to enjoy widespread adoption among practicing clinicians. This said, it remains one of the truly valid means by which to gain an appreciable understanding of children in their natural environment and remains a cornerstone for school psychologists and empirical investigations. A practical alternative is to video record children's behavior for later coding and analysis; however, this raises thorny ethical concerns and is unlikely to be permitted in most educational settings. Actometers represent a reasonable compromise between the rather limited and somewhat unreliable pedometer and highly precise but costly actigraph until the latter becomes more affordable for clinical practices. Either leg or wrist placement is preferred, coupled with a supplementary daily activity log to facilitate interpretation. Due to the limitations of the actometers, such a log must include the starting and ending "times" shown on the actometer for each activity listed in order to be useful. Multiday samples of behavior are recommended, particularly for children suspected of ADHD, owing to their high within- and across-day variability. Combining activity level data with information gleaned from extensive histories, semistructured clinical interview, and appropriate rating scales is required for effective case conceptualization, consistent with evidence-based practice.

FUTURE DIRECTIONS

Little is known about the activity level in children with clinical disorders relative to the burgeoning literature on normal development. Extant literature suggests that activity level differences associated with most clinical disorders of childhood represent quantitative deviations. For example, heightened activity level is frequently observed in children with attention-deficit/hyperactivity disorder (ADHD) and some forms of anxiety, whereas lower-than-normal levels of activity are often related to childhood depression. Differences in activity level topography—activity level's unique form and expression in a particular situation or setting—may also have special meaning for understanding the movement of some children. Topographical differences are frequently observed in children with autism (e.g., stereotypic, ritualistic behavior), who display self-injurious behavior, and as a result of untoward side effects associated with short-term (e.g., akathesia) and chronic (e.g., tardive dyskinesia) neuroleptic administration. Measuring activity level in children with particular clinical disorders using the measures reviewed must be supplemented by behavioral observation to understand the subtleties, contextual factors, and possible functional nature of children's motor activity. Actigraphs may play an increasingly larger role in clinical assessments, but they are functionally limited until standardized age and gender norms are developed.

REFERENCES AND RESOURCES

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Aman, M. G., & Werry, J. S. (1984). The revised behavior problem checklist in clinic attenders and nonattenders: Age and sex effects. *Journal of Clinical Child Psychology, 13,* 237–242.

Ambulatory Monitoring. (2004). The micromini motionlogger actigraph and family of single-sensor recorders [User's manual to accompany Act Millennium version 3.5.0.0 and higher (3rd ed.)]. New York: Author. Retrieved from http://www.ambulatory-monitoring.com/, 7 October 2004.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders, 3rd edition (DSM-III).* Washington, DC: Author.

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

Angello, L. M., Volpe, R. J., DiPerna, J. C., Gureasko-Moore, S. P., Gureasko-Moore, D. P., Nebrig, M. R., & Ota, K. (2003). Assessment of attention-deficit/hyperactivity disorder: An evaluation of six published rating scales. *School Psychology Review, 32,* 241–262.

Aronen, E. T., Fjällberg, M., Paavonen, E. J., & Soininen, M. (2002). Day-length associates with activity level in children living at 60 degrees north. *Child Psychiatry and Human Development, 32,* 217–226.

Barkley, R. A. (1988). Child behavior rating scales and checklists. In M. Rutter, A. H. Tuma, & I. S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 113–155). New York: Guilford Press.

Barkley, R. A., & Ullman, D. G. (1975). A comparison of objective measures of activity and distractibility in hyperactive and nonhyperactive children. *Journal of Abnormal Child Psychology, 3,* 231–244.

Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10,* 601–610.

Bell, R. Q. (1968). Adaptation of small wristwatches for mechanical recording of activity in infants and children. *Journal of Experimental Child Psychology, 6,* 302–305.

Butcher, J. E., & Eaton, W. O. (1989). Gross and fine motor proficiency in preschoolers: Relationships with free play behavior and activity level. *Journal of Human Movement Studies, 16,* 27–36.

Campbell, D. W., & Eaton, W. O. (1999). Sex differences in the activity level of infants. *Infant and Child Development, 8,* 1–17.

Conners, C. K., Parker, J. D. A., Sitarenios, G., & Epstein, J. N. (1998). The Revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 257–268.

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998). Revision and restandardization of the Conners' Teacher Rating Scale (CTRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 279–291.

Conoley, J. C., & Impara, J. C. (1995). *The twelfth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

DeFries, J. C., Gervais, M. C., & Thomas, E. A. (1978). Response to 30 generations of selection for open-field activity in laboratory mice. *Behavior Genetics, 8,* 3–13.

Demaray, M. K., Elting, J., & Schaefer, K. (2003). Assessment of attention-deficit/hyperactivity disorder (ADHD): A comparative evaluation of five, commonly used, published rating scales. *Psychology in the Schools, 40,* 341–361.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD rating scale—IV: Checklists, norms, and clinical interpretation.* New York: Guilford Press.

DuPaul, G. J., Rapport, M. D., & Perriello, L. M. (1991). Teacher ratings of academic skills: The development of the Academic Performance Rating Scale. *School Psychology Review, 20,* 284–300.

Eaton, W. O. (1983). Measuring activity level with actometers: Reliability, validity, and arm length. *Child Development, 54,* 720–726.

Eaton, W. O. (1994). Temperament, development, and the five-factor model: Lessons from activity level. In G. F. Habversow, G. A. Kohastamm, & R. P. Mecten (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 173–187). Hillsdale, NJ: Erlbaum.

Eaton, W. O., McKeen, N. A., & Campbell, D. W. (2001). The waxing and waning of movement: Implications for psychological development. *Developmental Review, 21,* 205–223.

Eaton, W. O., McKeen, N. A., & Saudino, K. J. (1996). Measuring human individual differences in general motor activity with actometers. In K. P. Ossenkopp, M. Kavaliers, & P. R. Sanberg (Eds.), *Measuring movement and locomotion: From invertebrates to humans* (pp. 79–92). Austin, TX: R. G. Landes.

Emerson, E. N., Crowley, S. L., & Merrell (1994). Convergent validity of the school social behavior scales with the Child Behavior Checklist and teacher's report form. *Journal of Psychoeducational Assessment, 12,* 372–380.

Fergusson, D. M., Lynskey, M. T., & Horwood, L. J. (1997). Attentional difficulties in middle childhood and psychosocial outcomes in young adulthood. *Journal of Child Psychology and Psychiatry, 38,* 633–644.

Gadow, K. D., & Sprafkin, J. (1997a). *ADHD Symptom Checklist—4 manual*. Stony Brook, NY: Checkmate Plus.

Gadow, K. D., & Sprafkin, J. (1997b). *Child symptom inventories norms manual*. Stony Brook, NY: Checkmate Plus.

Gadow, K. D., & Sprafkin, J. (2002). *Child Symptom Inventory 4: Screening and norms manual*. Stony Brook, NY: Checkmate Plus.

Gadow, K. D., Sprafkin, J., & Nolan, E. E. (1996). *Attention Deficit Hyperactivity Disorder School Observation Code*. Stony Brook, NY: Checkmate Plus Ltd.

Gadow, K. D., Sprafkin, J., Salisbury, H., Schneider, J., & Loney, J. (2004). Further validity evidence for the teacher version of the Child Symptom Inventory—4. *School Psychology Quarterly, 19*, 50–71.

Gavarry, O., Giacomoni, M., Bernard, T., Seymat, M., & Falgairette, G. (2003). Habitual physical activity in children and adolescents during school and free days. *Medicine & Science in Sports & Exercise, 35*(3), 525–531.

Gilliam, J. E. (1995). *Examiners manual for the Attention-Deficit/Hyperactivity Disorder Test: A method for identifying individuals with ADHD*. Austin, TX: Pro-Ed.

Halverson Jr., C. F., & Waldrop, M. F. (1973). The relations of mechanically recorded activity level to varieties of preschool play behavior. *Child Development, 44*, 678–681.

Hogan, A. E., Quay, H. C., Vaughn, S., & Shapiro, S. K. (1989). Revised Behavior Problem Checklist: Stability, prevalence, and incidence of behavior problems in kindergarten and first-grade children. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 103–111.

Holland, M. L., Gimpel, G. A., & Merrell, K. W. (1998). Innovations in assessing ADHD: Development, psychometric properties, and factor structure of the ADHD Symptoms Rating Scale (ADHD-SRS). *Journal of Psychopathology & Behavioral Assessment, 20*, 307–332.

Impara, J. C., & Plake, B. S. (1998). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Johnson, C. F. (1971). Hyperactivity and the machine, the actometer. *Child Development, 42*, 2105–2110.

Kramer, J. J., & Conoley, J. C. (1992). *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Lahey, B. B., & Piacentini, J. C. (1985). An evaluation of the Quay–Peterson Revised Behavior Problem Checklist. *Journal of School Psychology, 23*, 285–289.

Lambert, N., Hartsough, C., & Sandoval, J. (1990). *Manual for the Children's Attention and Adjustment Survey*. Circle Pines, MN: American Guidance Service.

Matier-Sharma, K., Perachio, N., Newcorn, J. H., Sharma, V., & Halperin, J. M. (1995). Differential diagnosis of ADHD: Are objective measures of attention, impulsivity, and activity level helpful? *Child Neuropsychology, 1*, 118–127.

McCarney, S. B. (1995a). *Attention deficit disorders evaluation scale* (2nd ed.). Columbia, MO: Hawthorne Educational Services.

McCarney, S. B. (1995b). *The Early Childhood Attention Deficit Disorders Evaluation Scale (ECADDES)*. Columbia, MO: Hawthorne Educational Services.

Milich, R., Widiger, T. A., & Landau, S. (1987). Differential diagnosis of attention deficit and conduct disorders using conditional probabilities. *Journal of Consulting and Clinical Psychology, 55*, 762–767.

Mitchell, J. V. (1985). *The ninth mental measurements yearbook—Volume II*. Lincoln, NE: Buros Institute of Mental Measurements.

Multimodal Treatment Study of Children with ADHD Cooperative Group (US). (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*, 1073–1086.

Neeper, R., Lahey, B. B., & Frick, P. J. (1990). *Manual for the Comprehensive Behavior Rating Scale for Children (CBRSC)*. San Antonio, TX: Psychological Corporation.

Partington, M. W., Lang, E., & Campbell, D. (1971). Motor activity in early life. I. Fries' congenital activity types. *Biology of the Neonate, 18,* 94–107.

Peach, W. J., & Cobb, S. E. (1990). Student ratings vs. teacher ratings on the Revised Behavior Problem Checklist. *Journal of Instructional Psychology, 17,* 194–196.

Plake, B. S., & Impara, J. C. (2001). *The fourteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Plake, B. S., Impara, J. C., & Spies, R. A. (2003). *The fifteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Porrino, L. J., Rapoport, J. L., Behar, D., Sceery, W., Ismond, D. R., & Bunney, W. E. (1983). A naturalistic assessment of the motor activity of hyperactive boys: I. Comparison with normal controls. *Archives of General Psychiatry, 40,* 681–687.

Quay, H. C. (1983). A dimensional approach to behavior disorder: The Revised Behavior Problem Checklist. *School Psychology Review, 12,* 244–249.

Quay, H. C., & Peterson, D. R. (1993). *The Revised Behavior Problem Checklist: Manual.* Odessa, FL: Psychological Assessment Resources.

Rapoport, J., Abramson, A., Alexander, D., & Lott, I. (1971). Playroom observations of hyperactive children on medication. *Journal of the American Academy of Child Psychiatry, 10,* 524–534.

Rapport, M. D. (1990). Controlled studies of the effects of psychostimulants on children's functioning in clinic and classroom settings. In C. K. Conners & M. Kinsbourne (Eds.), *Attention Deficit Hyperactivity Disorder* (pp. 77–111). Munich, Germany: Medizin Verlag Munchen.

Rapport, M. D., Murphy, H. A., & Bailey, J. S. (1982). Ritalin vs. response cost in the control of hyperactive children: A within-subject comparison. *Journal of Applied Behavior Analysis, 15,* 205–216.

Rapport, M. D., Scanlan, S. W., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *Journal of Child Psychology and Psychiatry, 40,* 1169–1183.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children: Manual.* Circle Pines, MN: American Guidance Service.

Routh, D. K., Schroeder, C. S., & O'Tuama, L. A. (1974). Development of activity level in children. *Developmenal Psychology, 10,* 163–168.

Saudino, K. J., & Eaton W. O. (1991). Infant temperament and genetics: An objective twin study of motor activity level. *Child Development, 62,* 1167–1174.

Schulman, J. L., & Reisman, J. M. (1959). An objective measure of hyperactivity. *American Journal of Mental Deficiency, 64,* 455–456.

Shelton, T. L., & Barkley, R. A. (1993). Assessment of attention-deficit/hyperactivity disorder in young children. In J. L. Culbertson & D. J. Willis (Eds.), *Testing young children: A reference guide for developmental, psychoeducational, and psychosocial assessments* (pp. 290–318). Austin, TX: Pro-Ed.

Sines, J. O. (1986). Normative data for the revised Missouri Children's Behavior Checklist—Parent form (MCBC-P). *Journal of Abnormal Child Psychology, 14,* 89–94.

Sleator, E. K., & Ullmann, R. K. (1981). Can the physician diagnose hyperactivity in the office? *Pediatrics, 67,* 13–17.

Spadafore, G. L., & Spadafore, S. J. (1997). *Spadafore attention deficit hyperactivity disorder rating scale.* Novato, CA: Academic Therapy Publications.

Sprafkin, J., Gadow, K. D., & Nolan, E. E. (2001). The utility of a DSM-IV-referenced screening instrument for attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders, 1,* 182–191.

Stevens, T. M., Kupst, M. J., Suran, B. G., & Schulman, J. L. (1978). Activity level: A comparison between actometer scores and observer ratings. *Journal of Abnormal Child Psychology, 6,* 163–173.

Swanson, J. M. (1992). *School-based assessments and interventions for ADD students.* Irvine, CA: K. C. Publications.

Teicher, M. H., Ito, Y., Glod, C. A., & Barber, N. I. (1996). Objective measurement of hyperactivity and attentional problems in ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 334–342.

Tryon, W. W. (1985). Human activity: A review of quantitative findings. In W. W. Tryon (Ed.), *Behavioral assessment in behavioral medicine* (pp. 257–299). New York: Springer.

Tryon, W. W. (1991). *Activity measurement in psychology and medicine.* New York: Plenum Press.

Tryon, W. W. (1993). Activity level and DSM-IV. In M. Hersen & S. M. Turner (Eds.), *Adult psychopathology and diagnosis* (4th ed., pp. 104–150). New York: Wiley.

Tryon, W. W., & Pinto, L. P. (1994). Comparing activity measurements and ratings. *Behavior Modification, 18,* 251–261.

Ullman, D. G., Barkley, R. A., & Brown, H. W. (1978). The behavioral symptoms of hyperkinetic children who successfully respond to stimulant drug treatment. *American Journal of Orthopsychiatry, 48,* 425–437.

Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1984). A new rating scale for diagnosing and monitoring of ADD children. *Psychopharmacology Bulletin, 20,* 160–164.

Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1997). *ACTeRS teacher and parent forms manual.* Champaign, IL: MetriTech, Inc.

van Egeren, L. A., Frank, S. J., & Paul, J. S. (1999). Daily behavior ratings among children and adolescent inpatients: The abbreviated child behavior rating form. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1417–1425.

Walters, C. E. (1965). Prediction of postnatal development from fetal activity. *Child Development, 36,* 801–808.

Werry, J. S. (1968). Developmental hyperactivity. *Pediatric Clinics of North America, 15,* 581–599.

Wilson, M. J., & Bullock, L. M. (1989). Psychometric characteristics of behavior rating scales: Definitions, problems, and solutions. *Behavioral Disorders, 14,* 186–200.

Zentall, S. S., & Leib, S. L. (1985). Structured tasks: Effects on activity and performance of hyperactive and comparison children. *Journal of Educational Research, 79,* 91–95.