Rating Scales in Attention-Deficit/Hyperactivity Disorder: Use in Assessment and Treatment Monitoring

C. Keith Conners, Ph.D.

Rating scales are valuable tools in both assessment and treatment monitoring. However, caution in their use is indicated because of several types of rater errors. Recent large-scale normative studies provide a set of instruments that cover child, adolescent, and adult ages, with separate gender norms and large representative samples. By including DSM-IV symptoms for ADHD in a proposed nationwide standardization of parent, teacher, and self-report scales, it is apparent that the proposed subtypes of ADHD are reasonable; however, item content in this standardization is somewhat broader than that proposed by DSM-IV. Empirical indexes were created and cross-validated, providing powerful discrimination between ADHD and non-ADHD samples. Separate scoring for the traditional DSM subtypes of ADHD allows both categorical and dimensional measures to be used in assessment and treatment monitoring. *(J Clin Psychiatry 1998;59[suppl 7]:24–30)*

In this paper I will not attempt a comprehensive review of rating scales in attention-deficit/hyperactivity disorder (ADHD), as several such reviews are available.[1–5] I will address some of the critical issues in the use of rating scales and describe recent developments that reflect changes in diagnostic criteria and social composition of normative samples. New scales for adolescent and adult ADHD will also be covered, and the particular uses of rating scales in psychopharmacologic trials will be highlighted.

ADVANTAGES OF RATING SCALES

In an earlier comprehensive review of rating scales, Conners and Barkley[3] described the rationale for rating scales as follows:

1. Ratings can draw upon a rater's often substantial previous experience with a child over long time intervals and diverse situations and circumstances;
2. They permit data to be gathered on rare and infrequent behaviors likely to be missed by in vivo assessments;
3. They are inexpensive to collect and extremely efficient in the time needed to gather information;
4. Many have normative data available for comparative purposes to show whether the drug brought the child's behavior closer to normal;
5. Numerous scales and checklists already exist, many having substantial information on their psychometric and practical properties;
6. Such ratings incorporate the opinions of "significant others" in the child's life whose ratings, regardless of accuracy or reliability, have substantial ecological importance as these are of the child's caregivers; and
7. Ratings permit the quantification of qualitative aspects of child behavior not readily gathered by other means.[3(p809)]

To this list one might also add the fact that, after three decades of childhood psychopharmacologic research, rating scales have proven to be drug-sensitive, with respect to both dose- and time-action phenomena.

DISADVANTAGES OF RATING SCALES

Rating scales have certain psychometric and practical advantages as noted above; but they also are subject to certain errors and forms of misuse and misinterpretation. First, there are particular errors that raters often make.[6] Of these, the *leniency error* refers to the tendency to judge the behavior too leniently with respect to its true severity or frequency; and the *severity error* is the opposite. There are also positive or negative *halo errors* in which the rater gives an unfair positive or negative slant to all items based

From the Department of Psychiatry, Duke University Medical Center, Durham, N.C.

Presented at the closed symposium "Current Issues in Attention Deficit Disorder," held November 13, 1996, Bloomingdale, Illinois. This supplement was sponsored by The Institute for Medical Studies, and both the meeting and the supplement were supported by an unrestricted educational grant from Wyeth-Ayerst Laboratories.

Reprint requests to: C. Keith Conners, Ph.D., Department of Psychiatry, Duke University Medical Center, Box 3431, Durham, NC 27710.

upon the judgment of one or more particular behaviors. For example, teachers will often rate children as deviant across a wide range of behaviors on the basis of their oppositional and disruptive behaviors, which happen to be particularly disturbing in the classroom, and inflate the degree of attentional or nonaggressive hyperactivity.[7–9] When raters make a *logical error*, they will give a rating at one time because they made a particular rating at another, which they feel must logically follow. Thus, a parent might say that the child is "restless and on the go" because he or she previously said the child was fidgety, despite the fact that these are quite separable behaviors. *Contrast errors* refer to the fact that ratings may vary depending upon whom the subject is compared with. When the parent contrasts the subject with a sibling rather than normal children of the same age, or when the parent uses a particular referent of what is "normal" a contrast error may occur. The prevalence rate of ADHD varies dramatically as a function of contrast phenomena.[10] The *recency error* refers to the tendency to rate the child according to the most recent episodes of behavior. Thus, parents or teachers may rate the child as much more deviant following a particularly bad episode in school or at home.

These errors have the effect of generating a certain amount of suspicion and distrust of rating scales among clinicians. However, the clinician is subject to these same errors of judgment. As in all skilled clinical investigation, an appraisal of the informant is essential, and a cross-checking of the information by using multiple sources and carefully evaluating the qualitative features of the information allow the clinician to place the information properly in context. Just as the marital therapist will carefully evaluate the veridicality of a disgruntled partner's negative judgments about the spouse, so must the annoyed teacher or parent be scrutinized for halo or severity errors. Thus, rating scale errors must always be minimized by judicious contextual evaluation of their meaning.

One might say, "But if it all comes down to clinical judgment, then why use rating scales in the first place?" First, because most rating scale information is veridical, useful, and accurately informative, despite the constant errors. Second, rating scales are not meant to stand alone, in either the diagnostic or treatment evaluation process. It is a perversion of their use to rely upon them solely as diagnostic or treatment outcome measures. Despite their limitations, they are, in general, reliable, valid, and treatment-sensitive.

NORMATIVE RATINGS AND SYMPTOM CRITERIA

The advent of the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association has created further interest in rating scales. Because the DSM spelled out diagnostic criteria for child psychiatric disorders in terms of specific lists of symptoms, questions have naturally arisen regarding the operational definition of symptom qualifiers such as "often" as well as the empirical meaning of "functional impairment." For example, what should count as, "*Often* fidgets with hands or feet or squirms in seat"? (one of the DSM-IV symptoms of ADHD). Typically, the answer to this question must be sought from parent, teacher, or other informants during the clinical assessment.

It is well known that observation of this behavior in the office can be quite misleading. Novelty and unfamiliarity of surroundings have a way of temporarily suppressing all but the most extreme childhood symptoms. It is the behavior in its typical context that is of interest to the clinician, not the brief and unrepresentative encounters in the clinic or office.

But what is a "normal" amount of fidgeting? Without a normative reference, it may be difficult to discern whether the observer or reporter of behavior is making a judgment that is too lenient or severe compared with assessment of behavior for the average child of that age, gender, and situation. Rating scales can ask this question of thousands of parents or teachers and determine statistically how the answers vary with age, gender, or other factors. Fidgeting, or leaving one's seat, like most childhood behaviors, falls along a continuum of frequency or severity, with no clear demarcation between an abnormal and a normal state. In certain cases of what used to be called the "driven hyperkinetic child," the behavior will be apparent in all situations. But for most children, one must rely upon the observer's intrinsic averaging process, whereby the individual in question is compared with one's experience of all children at the appropriate developmental stage.

It was a natural step for investigators to gather empirical normative data on the specific symptoms (items) from the DSM criteria lists, which themselves were often a rephrasing of items commonly found on rating scales. Thus, DSM was strongly influenced by the empirical experience with rating scales, and the latter in turn have been shaped by the specific behaviors chosen to represent symptomatic criteria in diagnostic algorithms. However, one of the difficulties with the frequent revisions of DSM has been that carefully developed norms on DSM items have been outdated by newer criteria. For example, Pelham et al.[11] collected teacher information on the DSM-III-R symptoms among a special education sample, as well as a large normative sample.[12] DuPaul et al.[13] also carried out a large study (729 children) of a DSM-III-R symptom list. As expected, the list produced restlessness/impulsivity and inattention factors, though inexplicably, restlessness was common to both factors. A major limitation of this study was the 18% return rate and the limited social-class and ethnic composition of the sample. A large body of reliability and validity data from 30 years of research has been summarized for the pre-DSM-IV versions of the Conners Scales.[14,15]

RECENT DEVELOPMENTS IN RATING SCALES

The ACQ Project

Rating scales developed as ad hoc responses to particular clinical and research needs, and as earlier reviews noted (see above), a wide variety of scales varying in length from a few items to several hundred items have been standardized. The standardization procedures, factor analytic methodologies, sample acquisition, and item selection have varied widely. While there was a great deal of overlap among different instruments, the differences in factor composition and variations due to local sampling factors were often substantial. It was this situation that led Tom Achenbach to seek funding from the American Psychological Foundation for a definitive national study. This study was to incorporate representative items from the three most widely used scales by Achenbach,[16] Conners,[17] and Quay[18]; items suggested by an international panel of experts; and a large, representative sample of parent raters from nonreferred and clinical sources. A census tract survey was conducted by a professional survey research team. The parent rating instrument (dubbed the ACQ) was administered to over 8500 normal and 8500 clinic parents. The results have been published.[5,19]

We had originally hypothesized 12 syndromes, based upon previous factor analytic studies, particularly those by the three senior investigators: aggression, anxious/depressed, ADD with and ADD without hyperactivity, delinquent, mean (in girls only), obsessive-compulsive, schizoid, sex problems, social ineptness, somatic complaints, and unresponsive/uncommunicative/withdrawn. Evidence for eight of the factors emerged from factor analyses: withdrawn, somatic, anxious/depressed, thought problems, social problems, attention problems, delinquent and aggressive. Of particular interest was that the expected factors of ADD with and without hyperactivity (based on then-current DSM-III concepts), did not appear. The "attention" factor included a mixture of immature behavior, poor concentration, impulsivity, restlessness, and poor school work.

We compared parent-reported problems and competencies for demographically matched nonreferred children and clinic-referred children using the ACQ Behavior Checklist, which includes 23 competence items, three competence scales, 216 problem items, the eight syndrome scales, internalizing, externalizing, and total competence and problem scores. Most items and scales discriminated significantly ($p < .01$) between referred and nonreferred samples. There were important sex and age differences in problem patterns, but regional and ethnic differences were minimal. Somewhat more problems and fewer competencies were reported for lower- than upper-socioeconomic-status children. Referral rates were similar in the most urban and rural areas, but they were significantly higher in areas of intermediate urbanization. Correlations of problem scores with those obtained 10 years earlier in a regional survey and with surveys in other countries showed considerable consistency in the rank order of prevalence rates among specific problems. Interview data from the survey sample yielded significantly higher ACQ problem scores for children who had fewer related adults in their homes; those who had more unrelated adults in their homes; those whose biological parents were unmarried, separated, or divorced; those whose families received public assistance; and those whose household or family members had received mental health services. Children who scored higher on externalizing than internalizing problems tended to have unmarried, separated, or divorced parents and to come from families receiving public assistance. However, there were greater proportions of both externalizing and internalizing patterns among children whose household or family members had received mental health services than among other children.

The ACQ made an important empirical contribution to the understanding of replicable child dimensional syndromes. The large representative normal sample and the large clinic sample ensured that potentially important demographic phenomena could be examined for the first time in a scientifically acceptable manner. The finding that regional and ethnic effects were minimal is particularly useful in view of the claims to the contrary based upon small, regionally biased samples previously available. However, there were important limitations of the research version of the scale: its length may have contributed to important variance due to respondent fatigue and the final version was too unwieldy for normal clinical use.

Restandardization of the Conners Scales

Although the original Conners scales continue to be used in a variety of clinical and research settings, the various scales were created prior to DSM criteria and were constructed and standardized on somewhat restrictive samples. The original teacher and parent scales were standardized on relatively small clinic and local Baltimore, Maryland, school samples. The revised versions[20] were standardized on a representative sample of census tract–identified families, but the original sample was restricted to urban Pittsburgh. The most popular scale, the brief 10-item subset known as the Hyperactivity Index, and the Teacher Rating Scale were standardized on over 10,000 normal schoolchildren.[21–23] Although large, this sample was the entire school population of but one Canadian city (Ottawa).

Moreover, this early work was characterized by item and factor analysis methodology of an earlier era. A particularly important limitation was the high intercorrelation of subscales caused by allowing items with high loadings on more than one factor. This work had the unfortunate effect of persuading some investigators that hyperactivity and conduct disorder dimensions were measuring the

same thing.[24] In order to counter this position, which was based upon a small sample of British clinic children, we employed 683 psychiatric and normal children (aged 4–15 years) to demonstrate that high intercorrelations of these scales are of purely methodological origin. When factor score coefficients were used to interpret factors, the hyperactive symptoms did not load on the conduct disorder factor. If factor scores were defined by the use of unit weights, then the intercorrelation between the hyperactive and conduct disorder factors was high. The use of factor score coefficients to define factors, on the other hand, produced uncorrelated factors. The results supported the idea that hyperactivity and conduct disorder are independent behavioral dimensions.[25*]

In the latest restandardization of the Conners Scales,[26] a consistent methodology was followed. Data for the child and adolescent versions of the scales were collected from over 200 sites in the United States and Canada (a total of about 11,000 ratings on approximately 5,000 children), selected to cover both urban and rural areas of all states and provinces, with a good representation of social class, ethnicity, age, and gender. A large initial pool of items was created based upon previous factor analytic studies by us and by others. The sample was split into a derivation sample and a cross-validation sample. The correlation matrix from the derivation sample was subjected to principal axis factoring, and a series of factor analyses was conducted to determine what items to retain. Items had to load significantly (> 0.30) on a given factor and load lower than 0.30 on the other factors. Following the rational approach to scale construction, an item was eliminated if it lacked conceptual coherence with its factor, and standard Scree test criteria (> 1.0) were applied to select the number of factors for rotation. In addition, we employed the split-half factor comparabilities method[27] to determine the most reliable factor solution. Finally, the factor analysis was repeated on the cross-validation sample using confirmatory factor analysis, and multiple criteria were used to assess the goodness-of-fit of the predicted model.

A further aim of the restandardization was to retain the useful properties of long and short scales. Long scales are particularly useful for initial characterization of a child, adolescent, or adult, when the broadest descriptive characterization is required, such as during an initial assessment. Shorter scales are useful for repeated measures, or when time availability for the raters is limited. The aim was to have equally powerful psychometric properties for the long and short scales so that no sacrifice in reliability or validity occurred between the different versions.

A further refinement of the new scales is the inclusion of DSM-IV subscales for hyperactivity/impulsivity and inattention, so that ADHD subtypes can be measured quantitatively and categorically. Those scales can be administered separately as an 18-item set, or as part of the long form. In the latter case, one would not only determine whether the child met criteria for ADHD, but also measure possibly important comorbidities.

In addition to the DSM-IV subscale, two other brief scales were developed. Perhaps the most useful of all the previous scales is the so-called Hyperactivity Index. This scale has generally outperformed other scales such as the Achenbach CBCL and Kendall and Wilcox's Self-Control Rating Scale in terms of differentiating ADD from psychiatric and normal controls.[28] This scale was originally the 10 items which were most highly loaded on each of the main factors from the long parent and teacher scales. Thus, it was more of a general psychopathology scale than a hyperactivity scale per se. This occasioned a certain amount of confusion among those who failed to read the fine print, as it were. Users complained that this scale was not unidimensional but complex, and therefore argued that it was misleading (in this author's opinion, a particularly misguided interpretation along these lines was presented by Ullman, Sleator, and Sprague[29]). A reanalysis of this scale using the large restandardization sample shows two clear factors, essentially internalizing and externalizing, or hyperactive-impulsive and emotional lability.[30] A large number of drug studies have relied upon this index as a key outcome measure because of its brevity (which allows frequent readministration), drug sensitivity, and ability to discriminate well-diagnosed ADHD from other samples.[31] It is undoubtedly the combination of externalizing and internalizing behavior that has made this instrument so useful in diagnostic and treatment studies. This index has been restandardized and retained, but a new index has been created using a different methodology.

A true ADHD index was created by using discriminant function analysis of carefully diagnosed ADHD children and adolescents compared with age and gender-matched subsamples of the standardization sample. This new 12-item index has high reliability and cross-validated sensitivity (98.2% and 97.1% in samples with N = 114 and N = 206 for the teacher scale; 92.3% and 100% in samples with N = 104 and N = 80 for the parent scale; and 90.7% and 90.7% in the adolescent self-report scale with N = 86 and N = 86). Specificity figures are of the same general magnitude (77% to 98%). Overall classification rates for both the initial and cross-validation samples range from 84% to 96%.

*When factor analyses are carried out with orthogonal rotation of factors, the factors are by definition as uncorrelated as the data will permit. If the exact weights that emerge from the analysis for each item are used to weight the items in creating factor scores, then this uncorrelated property will be retained. However, for convenience, one often simply sums the raw item scores to get total factor scores, which are then usually standardized (e.g., with t scores). To the extent that the same items are used in different factors, the factors will be correlated. If one's purpose is to provide an instrument with relatively independent factors, it behooves one to include items on a factor that have negligible loadings on all other factors.

A New Self-Report Adolescent Scale

Relatively little work has been done on adolescents with ADHD. As children enter middle school, reports of teachers become less satisfactory, and there are more areas of function that are likely to escape the perusal of parents. Although self-report is subject to the adolescent's proclivity to "fake good," it is generally accepted that self-report is useful as a complement to parent and teacher scales, particularly as regards internalizing states.[32] Moreover, there are important developmental differences that require a different set of items. We had previously developed an adolescent self-report scale to cover problems of concentration, restlessness, self-control, anger, social interactions, self-esteem, learning, mood states, and family problems.[33] As part of the restandardization project, this scale was standardized on 3486 adolescents between the ages of 12 and 17 years of age. By using the methodology described above, the original item set was reduced from 100 to 87 items for the long form, and 27 items for the short version. Six replicable factors emerged with good reliability and excellent sensitivity and specificity.[34,35]

In summary, Table 1 lists the child and adolescent scales, factor names, and number of items for the revised Conners Scales from the restandardization studies.

Table 1. Conners Rating Scales: Revised Child and Adolescent Scales (Long Versions)

	Items
Parent Scales[36]	
Factors	
Oppositional	10
Cognitive problems	12
Hyperactivity	9
Anxious/shy	8
Perfectionism	7
Social problems	5
Psychosomatic	6
Specialty Scales	
ADHD index	12
Global index: restless-impulsive	6
Global index: emotional lability	6
Global index: total score	12
DSM-IV inattentive	9
DSM-IV hyperactive-impulsive	9
DSM-IV total score	18
Teacher Scales[37]	
Factors	
Oppositional	6
Cognitive problems	7
Hyperactivity	7
Anxious/shy	6
Perfectionism	6
Social problems	5
Specialty Scales	
ADHD index	12
Global index: restless-impulsive	6
Global index: emotional lability	6
Global index: total score	12
DSM-IV inattentive	9
DSM-IV hyperactive-impulsive	9
DSM-IV total score	18
Adolescent Self-Report Scale[35]	
Factors	
Family problems	12
Emotionality	12
Conduct problems	12
Anger control problems	8
Hyperactivity	8
Specialty Scales	
ADHD index	12
DSM-IV inattentive	9
DSM-IV hyperactive-impulsive	9
DSM-IV total score	18

Adult ADHD Rating Scale

Several rating scales for adult ADHD are available. Spencer et al.[38] successfully used the DSM-III-R childhood symptom scale[4] in a methylphenidate treatment study with ADHD adults. Ward et al.[39] developed a 61-item adult scale using retrospective Conners Parent Ratings as a validation criterion and items taken from an earlier monograph.[40] Discriminating scale items were identified using a sample of 100 "normal" adults, 70 adult unipolar outpatients, and 81 adult outpatients with a diagnosis of adult ADHD. A factor analysis of 310 fathers and 305 mothers of children referred for ADHD revealed factors of conduct problems, learning problems, stress intolerance, attention problems, and poor social skills for males. A somewhat different factor structure emerged for females: dysphoria, impulsive/conduct, learning problems, attention, and organization problems. The restricted item sets (DSM-III-R and early concepts of minimal brain dysfunction [MBD]), as well as the small and idiosyncratic normative samples, render these scales problematic from a psychometric point of view, although their usefulness in drug trials is encouraging of the general concept of an adult ADHD scale.

We developed a new adult scale using methodology as described above (Conners C, Erhardt D, Epstein J, et al. 1997. Unpublished data). The derivation sample consisted of 840 normal adults between the ages of 18 and 81 years of age. The cross-validation sample consisted of 167 adults referred to an outpatient clinic for ADHD. A large pool of items was created based upon recent literature and recommendations by a number of clinicians and was ultimately reduced to a set of 42 items. The four replicable factors that emerged included inattention/memory problems, hyperactivity/restlessness, impulsivity/emotionality, and self-esteem. The scales show good reliability, sensitivity, and specificity.

Rating Scales in Drug Trials With Children

Rating scales have been used as subject selection and outcome measures since the 1960s. A large body of data has been collected on hundreds of drug trials. Several meta-analyses are in agreement in showing very large effect sizes of drug versus placebo comparisons using parent and teacher rating scales.[41–43] Scales such as the Hyperactivity Index show very good sensitivity to dose effects. However, careful analysis of individual cases reveals that

at least three different patterns of dose-response effect are evident. The most typical is a linear effect of increasing benefit with increasing dose. Less common is a "threshold effect," in which nothing happens with increasing dose until a given level, when suddenly improvement occurs. Finally, there is the classic quadratic effect, in which there is first an improvement with each dose, and then a worsening, sometimes to a level worse than the placebo phase.[44–46] These studies suggest the importance of individualizing each child's drug dose rather than relying upon preestablished criteria based upon group outcome studies.

The recently developed ADHD Index[47] for parents, teachers, and adolescents should provide a sensitive measure of drug changes on those symptoms that most clearly discriminate ADHD children from normals. Because of their brevity (12 items) and excellent psychometric properties, these indexes are well suited for repeated measures in clinical drug trials. Similarly, the DSM-IV scales now provide for the first time a standardized measure for the criterion symptoms of the DSM-IV definition of ADHD. When a Food and Drug Administration (FDA) drug trial explicitly targets ADHD, it would seem important to include both the categorical and dimensional measures of the concept.

Clinical global rating scales typically try to capture an overall severity or improvement response, using judgments by clinicians with access to other available data. The most commonly used format is the National Institute of Mental Health (NIMH) version.[48] These global ratings serve an important function in providing a single overall endpoint for controlled trials. However, global scales are sometimes quite unreliable. In a recent exercise with over 75 clinical investigators and trial coordinators, this author presented parent, teacher, and interview data to the clinicians and had them rate severity and improvement as one might typically do with such data in a clinical trial. The initial response showed wide disparities among investigators. With repeated practice using new cases, rater concordance gradually developed. It was apparent that some investigators placed greatest weight on parent information, while others relied more upon teacher or interview data in forming their judgments. The purpose of global instruments is not to produce stereotypic judgments based largely on one data source, but to synthesize data in a manner not possible from single sources. However, unless careful reliability training occurs, there is great potential for a high noise level, which could render a trial invalid by creating a type II error (not detecting a true difference that exists between drug and placebo).

Typically, drug trials with newer agents will need to ascertain the impact of the drug on other targets besides just those most associated with ADHD. For example, anxiety, depression, perfectionism, and somatic complaints as well as ADHD symptoms may also increase or decrease with certain dosages or drugs. In those cases it seems judicious to use the longer scales as baseline and end-of-treatment measures so that all dimensions of behavior are covered.

There is little experience with symptom changes due to drugs in adolescent trials involving self-report. However, one study found drug effects from self-report similar to those from parent and teacher report; but different patterns of side effects were obtained from the self- and other-reports.[13] Since adolescents may be better reporters of internal states than either parents or teachers, it would seem to be useful to include self-report measures in trials with adolescents.

SUMMARY

As we have stated in prior reviews,[3] there is no single rating instrument that meets the needs of all investigators or clinicians. With the current generation of new instruments, there is both greater psychometric sophistication and more representative normative and clinical samples than were available in the past. Of particular interest is the fact that relatively brief forms are now available that retain the psychometric power of much longer and more cumbersome forms. This minimizes subject burden and user time.

On the other hand, given the current state of child psychiatric nosology, there are occasions when broader coverage is important because of the ever-present comorbidities that affect both subject selection and response to treatment. Rating scales offer important information on dimensionalities of behavior that have been well-established by decades of empirical work. But these data should always be seen in the context of a number of clinical operations such as interview, history, and other kinds of documentation. They are not meant to stand on their own or to relieve clinicians of the burden of careful synthesis of all information and the informed use of intuition and clinical judgment.

REFERENCES

1. Conners CK. Rating scales for use in drug studies with children: special issue on children. Psychopharmacol Bull 1973:24–42
2. Conners CK. Rating scales. In: Noshpitz J, ed. Handbook of Child Psychiatry, vol 1. New York, NY: Basic Books; 1979:675–691
3. Conners CK, Barkley RA. Rating scales and checklists for child psychopharmacology. Psychopharmacol Bull 1985;21:809–838
4. Barkley RA. Attention Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment. New York, NY: Guilford Press; 1990
5. Achenbach TM, Howell CT, Quay HC, et al. National survey of problems and competencies among four- to sixteen-year-olds: parents' reports for normative and clinical samples. Monogr Soc Res Child Dev 1991;56 (suppl 3):1–131
6. Aiken LR. Rating Scales and Checklists: Evaluating Behavior, Personality, and Attitude. New York, NY: John Wiley & Sons; 1996
7. Abikoff H, Courtney M, Pelham WEJ, et al. Teachers' ratings of disruptive behaviors: the influence of halo effects. J Abnorm Child Psychol 1993;21 (suppl 5):519–533
8. Schachar RJ, Sandberg S, Rutter M. Agreement between teachers' ratings and observations of hyperactivity, inattentiveness, and defiance. J Abnorm Child Psychol 1986;14(suppl 2):331–345
9. Siegel LJ, Dragovich SL, Marholin DI. The effects of biasing information on behavioral observations and rating scales. J Abnorm Child Psychol 1976;4(suppl 3):221–233

10. Holborow PL, Berry P, Elkins J. Prevalence of hyperkinesis: a comparison of three rating scales. Journal of Learning Disabilities 1984;17(suppl 7): 411–417
11. Pelham WE, Evans SW, Gnagy EW, et al. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders: prevalence, factor analyses, and conditional probabilities in a special education sample. School Psychology Review 1992;21;285–299
12. Pelham WE, Gnagy EM, Greenslade KE, et al. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. J Am Acad Child Adolesc Psychiatry 1992;31:210–218
13. DuPaul GJ. Parent and teacher ratings of ADHD symptoms: psychometric properties in a community-based sample. Special issue: child psychopharmacology. Journal of Clinical Child Psychology 1991;20:245–253
14. Conners CK. The Conners Rating Scales: use in clinical assessment, treating planning and research. In: Maruish M, ed. Use of Psychological Testing for Treatment Planning and Outcome Assessment. Hillsdale, NJ: L Erlbaum Associates; 1994
15. Wainright A. Conners' Rating Scales: 30 Years of Research. Toronto, Canada: Multi-Health Systems; 1996
16. Achenbach T, Edelbrock C. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington: University of Vermont, Department of Psychiatry; 1983
17. Conners CK. Manual for Conners' Rating Scales. North Tanawanda, NY: Multi-Health Systems; 1989
18. Quay HC. Measuring dimensions of deviant behavior: the behavior problem checklist. J Abnorm Child Psychol 1977;5;277–287
19. Achenbach TM, Conners CK, Quay HC, et al. Replication of empirically derived syndromes as a basis for taxonomy of child/adolescent psychopathology. J Abnorm Child Psychol 1989;17(suppl 3):299–323
20. Goyette CH, Conners CK, Ulrich RF. Normative data on revised Conners Parent and Teacher Rating Scales. J Abnorm Child Psychol 1978;6(suppl 2):221–236
21. Trites R, Blouin A, Ferguson H, et al. The Conners' teacher rating scale: an epidemiologic inter-rater reliability and follow-up investigation. In: Gadow K, Loney J, eds. Psychosocial Aspects of Drug Treatment for Hyperactivity. Boulder, Co: Westview Press; 1981
22. Trites RL, Blouin AG, Laprade K. Factor analysis of the Conners Teacher Rating Scale based on a large normative sample. J Consult Clin Psychol 1982;50(suppl 5):615–623
23. Trites RL, Dugas E, Lynch G, et al. Prevalence of hyperactivity. J Pediatr Psychol 1979;4(suppl 2):179–188
24. Sandberg ST, Rutter M, Taylor E. Hyperkinetic disorder in psychiatric clinic attenders. Dev Med Child Neurol 1978;20(suppl 3):279–299
25. Blouin AG, Conners CK, Seidel WT, et al. The independence of hyperactivity from conduct disorder: methodological considerations. Can J Psychiatry 1989;34(suppl 4):279–282
26. Conners CK. The Conners Rating Scales (Revised): Technical Manual. Toronto, Canada: 1997
27. Everett J. Factor comparability as a means of determining the number of factors and their rotation. Multivariate Behavioral Research 1983;18: 197–218
28. Zelko FA. Comparison of parent-completed behavior rating scales: differentiating boys with ADD from psychiatric and normal controls. J Dev Behav Pediatr 1991;12(suppl 1):31–37
29. Ullmann RK, Sleator EK, Sprague RL. A change in mind: the Conners Abbreviated Rating Scales reconsidered. J Abnorm Child Psychol 1985;13 (suppl 4):553–565
30. Parker JDA, Sitarenios G, Conners CK. Abbreviated Conners' Rating Scales revisited: a confirmatory factor analytic study. Journal of Attention Disorders 1996;1:55–62
31. Conners CK. The Conners Rating Scales: use in clinical assessment, treatment planning and research. In: Maruish M, ed. Use of Psychological Testing for Treatment Planning and Outcome Assessment. Hillsdale, NJ: Lawrence Erlbaum Associates; 1994
32. Klorman R. Some thoughts on the diagnosis of ADD/H in adolescence. Psychopharmacol Bull 1985;21:913–914
33. Conners CK, Wells KC. ADD-H adolescent self-report scale. Psychopharmacol Bull 1985;21:921–922
34. Robin AL. Adolescent self-report of ADHD symptoms: the ADHD reports. 1994;2:4–6
35. Conners CK, Wells KC, Parker JDA, et al. A new self-report scale for assessment of adolescent psychopathology: factor structure, reliability, validity, and diagnostic sensitivity. J Abnorm Child Psychol. In press
36. Conners CK, Sitarenios G, Parker JDA, et al. The Revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. J Abnorm Child Psychol. In press
37. Conners CK, Sitarenios G, Parker JDA, et al. The Revised Conners' Teacher Rating Scale: factor structure, reliability, and criterion validity. J Abnorm Child Psychol. In press
38. Spencer T, Wilens T, Biederman J, et al. A double-blind, crossover comparison of methylphenidate and placebo in adults with childhood-onset attention-deficit hyperactivity disorder. Arch Gen Pyschiatry 1995;52: 434–443
39. Ward MF, Wender MF, Reimherr FW. The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. Am J Psychiatry 1993;150:885–890
40. Wender PH. Minimal Brain Dysfunction in Children. New York, NY: John Wiley & Sons; 1971
41. Kavale K. The efficacy of stimulant drug treatment for hyperactivity: a meta-analysis. Journal of Learning Disabilities 1982;15:280–289
42. Ottenbacher JJ, Cooper HM. Drug treatment of hyperactivity in children. Dev Med Child Neurol 1983;25:356–358
43. Thurber S, Walker CE. Medication and hyperactivity: a meta-analysis. J Gen Psychol 1983;108:79–86
44. Rapport M, Stoner G, DuPaul G, et al. Attention deficit disorder and methylphenidate: a multilevel analysis of dose-response effects on children's impulsivity across settings. J Am Acad Child Adolesc Psychiatry 1989;27: 60–69
45. Rapport MD. Controlled studies of the effects of psychostimulants on children's functioning in clinic and classroom. In: Conners CK, Kinsbourne M, eds. ADHD: Attention Deficit Disorder. Munich, Germany: MMV Medizin Verlag; 1990
46. Sprague RL, Sleator EK. Methylphenidate in hyperkinetic children: differences in dose effects on learning and social behavior. Science 1977;198: 1274–1276
47. Conners CK, Conners' Rating Scales: Revised Technical Manual. North Tanawanda, NY: Multi-Health Systems; 1997
48. National Institute of Mental Health. Special issue on rating scales and assessment instruments for use in pediatric psychopharmacology research. Psychopharmacol Bull 1985;21(4)