Published in final edited form as:

J Atten Disord. 2016 August ; 20(8): 665–673. doi:10.1177/1087054712475083.

Agreement Among Categorical, Dimensional, and Impairment Criteria for ADHD and Common Comorbidities

Joyce Sprafkin¹, Elizabeth Steinberg², Kenneth D. Gadow¹, and Deborah A. G. Drabick²

¹Stony Brook University, NY, USA

²Temple University, Philadelphia, PA, USA

Abstract

Objective—To compare the results of categorically based versus dimensionally based scoring algorithms for a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*)–referenced teacher rating scale for assessing ADHD and commonly co-occurring conditions and to determine their relative agreement with ratings of symptom-induced impairment.

Method—Teachers completed Child and Adolescent Symptom Inventory–4R (CASI-4R) ratings for 1,092 youth (ages 6–18 years) referred to a child and adolescent psychiatry outpatient service. Caseness was determined according to *DSM-IV* symptom count (categorical model) and *T*-score (dimensional model) criteria.

Results—Agreement between symptom count and *T*-score cutoffs was generally good (kappa 0.61) for ADHD-Inattentive, ADHD-Hyperactive-Impulsive, ADHD-Combined (except adolescent females), Oppositional Defiant Disorder, and Conduct Disorder, but this was not the case for anxiety and depressive disorders where only 15% of kappas were good. Agreement of impairment cutoff with *T*-score and symptom count cutoffs ranged from poor to good.

Conclusion—In general, although in many cases CASI-4R categorical and dimensional scoring algorithms generated similar results, there was considerable variability across disorders, age groups, scoring method, and in some cases, gender. Moreover, symptom counts and *T*-scores are not a proxy for assessing impairment suggesting that each scoring strategy likely provides unique information for clinical decision-making.

Keywords

ADHD; DSM-IV; rating scales; functional impairment

The use of behavior rating scales in the assessment of ADHD and commonly co-occurring conditions has become standard practice in many Western countries. The vast majority of

Corresponding Author: Joyce Sprafkin, Department of Psychiatry and Behavioral Sciences, Stony Brook University, Stony Brook, NY 11794-8790, USA, joyce.sprafkin@stonybrook.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Financial Disclosures

Joyce Sprafkin & Kenneth D. Gadow: shareholders, Checkmate Plus, publisher of Child and Adolescent Symptom Inventory; Elizabeth Steinberg & Deborah A. G. Drabick: none.

such rating scales can be classified as reflecting either categorical or dimensional models of psychopathology. However, this distinction is greatly muddled by the facts that (a) the symptoms of a categorical disease can be scored on a Likert-type scale to generate a “dimensional” measure of symptom severity and (b) dimensions derived from factor analyses are typically converted into a categorical measure with the aid of cutoff scores (e.g., *T*-scores 1.5 or 2.0 standard deviations above the mean indicate some level of probable disorder). As there are a number of excellent reviews of the history, strengths, and weaknesses of each model (e.g., Coghill & Sonuga-Barke, 2012), this topic is not discussed here.

In the United States, the archetypal example of the categorical model is the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association [APA], 2000) where structured or semistructured interviews (e.g., Orvaschel, Puig-Antich, Chambers, Tabrizi, & Johnson, 1982; Reich, Shayka, & Taibleson, 1991) were initially the most widely used assessment instruments derived from the model. At a minimum, these interviews typically include the symptoms of the disorder, a cutoff score based on the prerequisite number of symptoms for a diagnosis (symptom count), and some type of impairment assessment. Owing to their seeming impracticality for everyday clinical evaluation (e.g., special training for the interviewer, time-consuming to administer), rating scales based on *DSM* symptoms rapidly gained popularity (e.g., Boyle et al., 1993; Gadow & Sprafkin, 1986, 2002; Pelham, Fabiano, & Massetti, 2005; Pelham, Gnagy, Greenslade, & Milich, 1992; Wolraich, Feurer, Hannah, Baumgaertel, & Pinnock, 1998). Given the omnipresence of the *DSM* system in the mental health field, some popular factor-analytically derived broadband rating scales also were modified to include *DSM-based* subscales and scoring procedures (Achenbach & Rescorla, 2000; Conners, 2002). Because there are many practical situations in which it is useful to compare clinically referred and typically developing youth, this eventually led to the generation of population-based norms for several *DSM*-oriented rating scales (e.g., Gadow & Sprafkin, 2002, 2008), from which cutoffs based on symptom severity could be determined. Nevertheless, there is relatively little research comparing the results obtained when using categorical versus dimensional scoring procedures for *DSM*-oriented rating scales.

One of the most important modifications made in the transition from *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., revised; *DSM-III-R*, APA, 1987) to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; APA, 1994) was the explicit inclusion of symptom-induced functional impairment as a diagnostic criterion for the presence of each disorder (i.e., included with the list of symptoms and other diagnostic criteria). One of the compelling reasons for this addition was that symptom criteria alone produce unrealistically high prevalence rates in epidemiological studies (e.g., Canino et al., 2004; Faraone, Sergeant, Gillberg, & Biederman, 2003; Healey, Miller, Castelli, Marks, & Halperin, 2008). Although there have been major reviews about research concerning functional impairment and its role in the assessment of children and adolescents (Bird et al., 2000; Gordon et al., 2006; Kazdin & Kagan, 1994; Rapee, Bogels, van der Sluis, Craske, & Ollendick, 2012; Wakefield, 1997; Winters, Collett, & Myers, 2005), there is a consensus that further research is needed owing to its importance in influencing service delivery and treatment planning.

The general aim of the present study was to compare symptom count, *T*-score, and impairment cutoff criteria for ADHD and commonly co-occurring conditions. Specifically, we compared the results of teachers' ratings using three different categorical scoring algorithms: prerequisite number of *DSM-IV*-specified symptoms (symptom cutoff), elevated *T*-scores based on a dimensional model of symptom severity (*T*-score cutoff), and symptom-induced impairment (impairment cutoff) in a large sample of clinically referred youth. To the best of our knowledge, this is one of the few studies to examine the degree of agreement (kappa) between these particular methods of parsing youth into clinical phenotypes and whether the magnitude of association varies by disorder. Prior research indicating that a semistructured psychiatric interview and a *DSM-IV-referenced* ADHD rating scale evidence satisfactory agreement (e.g., Sprafkin, Gadow, Salisbury, Schneider, & Loney, 2002) suggests that the two approaches generate similar results. However, for conditions such as Generalized Anxiety Disorder (GED) that require specific symptoms to be present (instead of giving equal weighting to all symptoms), agreement may be less strong. Furthermore, because several studies have reported modest to moderate agreement between symptom and impairment severity (Gathje, Lewandowski, & Gordon, 2008; Gordon et al., 2006; Rapee et al., 2012), we predicted this would also be the case for *T*-score and impairment cutoffs. In everyday clinical settings, impairment is implicit in help-seeking behavior; nevertheless, the relation of impairment cutoff with *DSM-IV* symptom and *T*-score cutoffs based on data from the same assessment instrument is relatively unstudied (Gadow, Kaat, & Lecavalier, 2012). Although it has been shown that in some instances teachers' reports are better predictors of correlates, impairment, or clinical course compared with parents' ratings (Drabick, Gadow, & Loney, 2007), relatively few studies have compared categorical versus dimensional scoring algorithms for teacher-completed ADHD rating scales, particularly with regard to associations with impairment. Therefore, the present study focuses exclusively on the teacher.

Method

Participants

Participants were teachers of 1,092 youth (763 males, 329 females) aged 6 to 18 years ($M = 12.1$, $SD = 3.4$) who were consecutive referrals (2004–2010) to a university hospital child psychiatry outpatient service. All youth were included in the study sample if they were within the specified age range, had IQs ≥ 70 (range = 70–144), and were not diagnosed as having autism spectrum disorder or schizophrenia spectrum disorder. The sample was divided into a younger (6–11 year olds; $n = 516$) and older (12–18 year olds; $n = 576$) cohort. Self-identified ethnicities were as follows: European American (79%), African American (7%), Hispanic American (10%), Native American (1%), Asian American (2%), and Other (1%). The two most common clinician-assigned diagnoses were ADHD and Oppositional Defiant Disorder (ODD); many youth with ODD were also comorbid for ADHD (30%). The study sample is described in greater detail in prior publications (e.g., Drabick & Gadow, 2012; Gadow & Drabick, 2012; Gadow, Kaat, et al., 2012). This retrospective chart review study was approved by a university Institutional Review Board, and appropriate measures were taken to protect patient (and rater) confidentiality.

Measures

DSM-IV symptoms—Teachers rated youth's symptoms using the Child and Adolescent Symptom Inventory–4R (CASI-4R; Gadow & Sprafkin, 2005). The CASI-4R is a 109-item teacher-completed behavior rating scale for evaluating youth 5 to 18 years old and combines the symptom modules from the Child Symptom Inventory–4 (Gadow & Sprafkin, 1986, 1994, 2002) and the Adolescent Symptom Inventory–4 (Gadow & Sprafkin, 1995, 2008). Prior research has documented satisfactory reliability (test–retest, interrater) and validity (association with relevant rating scales, structured interviews, and clinician diagnoses) of the teacher-completed Child Symptom Inventory–4 (e.g., Gadow & Sprafkin, 2011; Gadow, Sprafkin, Salisbury, Schneider, & Loney, 2004; Mattison, Gadow, Sprafkin, Nolan, & Schneider, 2003). Individual items bear one-to-one correspondence with *DSM-IV* (APA, 1994) symptoms (i.e., high content validity) and are rated as *never*, *sometimes*, *often*, or *very often*. We considered the following scales (α s) based on their relevance to evaluations of ADHD and its common comorbid disorders: ADHD, Inattentive type (ADHD-I; $\alpha = .93$); ADHD, Hyperactive-Impulsive type (ADHD-HI; $\alpha = .95$); ODD ($\alpha = .95$); Conduct Disorder (CD; $\alpha = .83$); GAD ($\alpha = .75$); Social Phobia ($\alpha = .85$); and Dysthymic Disorder ($\alpha = .81$). Following the set of symptom-based items for each disorder, teachers were asked, “How often do the behaviors in this category interfere with the student’s social or academic functioning?” Using the scoring method described in Gadow et al. (2010), responses of *never* and *sometimes* are coded as *not impaired* and those of *often* and *very often* are coded as *impaired*. The ADHD symptom category contains one impairment rating for all 18 symptoms.

Using the categorical scoring method, for the majority of items, symptoms rated as occurring *often* and *very often* are considered clinically significant and those rated *never* and *sometimes* are not. Youth receive a symptom cutoff score if the number of clinically significant symptoms is equal to or greater than the number of symptoms specified in *DSM-IV* as being necessary for a diagnosis.

Using the dimensional scoring method, responses are scored as never (0), sometimes (1), often (2), or very often (3). Item scores are summed to generate a symptom severity score for each category. Symptom severity scores are then compared to the scores from the normative samples used for the Child Symptom Inventory Teacher Checklist (for 5–12 year olds) and the Adolescent Symptom Inventory Teacher Checklist (for 12–18 year olds). Sample sizes for the normative samples were 1,323 children (662 males, 661 females) and 994 adolescents (508 males, 486 females). More detailed information about the normative samples is contained elsewhere (Gadow & Sprafkin, 2002, 2008).

The cutoffs examined in this study were symptom count cutoff scores (i.e., categorical method) and *T*-scores of 1½ and 2 standard deviations above the mean (*T*-scores 65 and 70, respectively; symptom severity or dimensional method). These cutoff scores were selected because they are among the most commonly used in the extant literature to differentiate individuals with and without problematic symptoms. A 1.5 *SD* cutoff score is often used as a “screening” criterion for identifying children who may be in need of a more in-depth mental health evaluation. This can be particularly helpful in identifying youth who

do not meet full symptom criteria for a specific disorder, but nevertheless evidence social or academic impairment attributed to the symptoms. A cutoff score ≥ 2 SDs is generally considered to be useful criterion for youth who are in likely need of mental health services.

Procedure

Prior to scheduling their initial clinic evaluation, parents were mailed a packet of materials including behavior ratings scales for both parent and teacher (e.g., CASI-4R), background information questionnaire, and permission for release of school, psychoeducational, and special education evaluation records. Teacher CASI-4R ratings were given to the school by parents, completed by teachers, and mailed directly to the clinic prior to the evaluation. Although both parents and teachers completed these measures for all children and adolescents who were evaluated regardless of their eventual diagnosis, analyses for the present study are limited to teachers' ratings. Procedures for evaluating patients are described in detail elsewhere (Gadow & Sprafkin, 2002, 2008).

Statistical Analyses

The kappa statistic was used to evaluate the extent of agreement in diagnostic classification (above vs. below cutoff) between symptom count and symptom severity cutoff scores for each of eight disorder categories from the CASI-4R Teacher Checklist. Separate calculations were performed for T -score cutoffs ≥ 65 and ≥ 70 (dimensional method) and symptom count cutoffs (categorical method). Results are presented by gender and age group (children/adolescents). In interpreting kappa, we used the criteria proposed by Altman (1997) and Kraemer, Kupfer, Clarke, Narrow, and Regier (2012) in which scores less than 0.20 are considered poor, scores between 0.21 and 0.40 are considered fair or acceptable, scores between 0.41 and 0.60 are considered moderate or realistic, scores between 0.61 and 0.80 are considered good, and scores above 0.80 are considered very good.

Given that the kappa statistic is influenced by base rates and the symmetry of the marginal totals (Feinstein & Cicchetti, 1990; Viera & Garrett, 2005), we also calculated the indices of positive agreement (p_{pos}) and negative agreement (p_{neg}), as recommended by Cicchetti and Feinstein (1990). These statistics elucidate the extent of agreement between the two scoring methods for decisions about the presence (p_{pos}) and absence (p_{neg}) of symptoms of each disorder (available upon request from author).

Because research based on this sample (Drabick & Gadow, 2012; Gadow, Kaat, et al., 2012) and other study samples (Gadow & Sprafkin, 2002, 2008) indicate differences between males and females as well as younger and older youth for a few symptom domains, analyses were conducted separately for these subgroups of patients.

Results

Overlap Between Categorical and Dimensional Methods

The kappa statistic indicating the degree of relationship between the categorical and dimensional cutoff methods is presented in Table 1. For ADHD-I and ADHD-HI ratings of the younger cohort, 7 out of 8 kappas were at least good (≥ 0.61) between symptom count

cutoffs and *T*-score cutoffs of 65 and 70. For the older cohort, the degree of agreement between the two methods was somewhat lower; 5 out of 8 kappas were 0.61. Only 2 out of the 8 kappas for ADHD-Combined (ADHD-C) were good. However, the *DSM* symptom count method (categorical) is based on two criteria (both ADHD-I and ADHD-HI symptom counts have to surpass a set criterion) whereas the *T*-score cutoff method is based on the distribution of total scores from the 18 ADHD items. Given the possibility that the modest correspondence between the categorical and dimensional scoring methods for ADHD-C was due to an imbalanced comparison, ADHD-C was rescored for the *T*-score method such that to meet criteria for ADHD-C, both the ADHD-I and ADHD-HI severity scores had to meet the designated *T*-score cutoff (either 65 or 70). Using this more parallel scoring approach, with the exception of adolescent females, all of the comparisons were good. For ODD, the degree of agreement between symptom count cutoffs and *T*-score cutoffs was very good for both younger (range = 0.68–0.93) and older (range = 0.60–0.93) youth. For CD, symptom count cutoffs yielded similar findings to *T*-score cutoffs for younger (0.78 and 0.88, for *T*-scores of 65 and 70, respectively) and older males (0.87 and 0.83). In contrast, there was only fair agreement between symptom count cutoffs and *T*-score cutoffs for CD symptoms among females of both age groups.

In contrast to the rather robust agreement between the results using symptom count cutoffs and *T*-score cutoffs for the disruptive behavior disorders, there was only poor to moderate agreement for the GAD, Dysthymic Disorder, and Social Phobia categories. The outcomes of the two scoring methods were particularly discrepant for ratings of GAD symptoms among older youth (range = 0.13–0.26) and only slightly better among younger youth (range = 0.37–0.54). For Dysthymic Disorder symptoms, in all but one case, agreement between the two scoring methods was in the fair to moderate range (0.22–0.54) for both age groups. Agreement between the two scoring methods was somewhat better for Social Phobia symptoms than for GAD and Dysthymic Disorder symptoms, with kappas ranging between 0.56 and 0.62 in the younger cohort. (Normative data were not available for older youth because there were too few items to derive a meaningful distribution.)

With regard to the indices of positive and negative agreement, when the kappa statistic is at least 0.70, the index of positive agreement was always at least 75%, and the index of negative agreement was always at least 85%. In other words, the two methods produce good and fairly even agreements about the presence and absence of symptom criteria. However, for lower kappa values, there were many instances where there was a large discrepancy between the indices of positive and negative agreement, and for the vast majority of these, the index of negative agreement was much larger than the index of positive agreement. Alternatively stated, when the two scoring methods yield fairly different results (low kappa), there was a tendency for the agreement to be much better for decisions about a youth falling below compared with above the cutoff criterion.

Association Between Impairment Ratings and Categorical and Dimensional Scoring

Tables 2 and 3 present the kappa statistic for the association between disorder category impairment ratings and both *T*-score cutoffs (65, 70) and symptom count cutoffs. Among younger youth (Table 2), 89% of the kappas between impairment ratings and both symptom

count and symptom severity ratings were at least good ($r = 0.61$) for ADHD-HI, ODD, and CD. The degree of association was in the moderate/realistic range (0.41–0.60) for GAD, Social Phobia, and ADHD-C and in the fair range (0.21–0.40) for ADHD-I. For ADHD-C, the association between impairment cutoff and *DSM-IV* symptom count cutoff was somewhat better for children (males = 0.61, females = 0.52) than adolescents (Table 3: 0.28 and 0.16, respectively). The only disorder category for which there was a clear advantage for one of the cutoff methods was Dysthymic Disorder; the symptom count cutoff was more closely associated with impairment ratings (0.59–0.74) than the *T*-score cutoff (0.26–0.37).

Among older youth (Table 3), the relationship between impairment ratings and symptom cutoff scores were most consistently in the good range for ODD (both symptom count and symptom severity). With only three exceptions, both symptom count and symptom severity cutoffs were moderately (0.41–0.60) related to impairment cutoffs for GAD, Dysthymic Disorder, and CD. For ADHD-I, impairment cutoffs corresponded better to *DSM-IV* symptom count (males = 0.57, females = 0.69) than to *T*-score (males = 0.32–0.37, females = 0.25–0.32) cutoffs. In contrast, for ADHD-HI and ADHD-C, the *T*-score cutoff method yielded greater agreement with impairment ratings than the *DSM-IV* symptom count method, which is consistent with the research literature indicating that the symptom count criteria set by *DSM-IV* are too high for adolescents (Ramtekkar, Reiersen, Todorov, & Todd, 2010).

Discussion

The results of this study support the integration of categorical and dimensional conceptual models of psychopathology in behavior rating scales of ADHD and co-occurring conditions. Specifically, cutoffs based on *T*-scores and *DSM-IV* symptom counts yield quite similar results for ratings of ADHD-I, ADHD-HI, ADHD-C (except adolescent females) and ODD among males and females. In both age groups, the majority of kappa values were good. Similarly, for CD, *T*-score and *DSM* symptom count cutoffs produced comparable findings among males but somewhat less so for females. One possible explanation for this discrepancy is that females referred for clinic evaluation are differentially more similar to males in terms of displaying antisocial behaviors (e.g., Gadow & Sprafkin, 2002); therefore, clinic girls may easily exceed *T*-score cutoffs (derived from norm-based comparisons) relative to *DSM* cutoffs. It is interesting to speculate about why there was far less overlap between the two scoring methods for ADHD-C in adolescent females than for the other cohorts. It is well established (e.g., Ramtekkar et al., 2010) that the *DSM-IV* ADHD symptom count criterion of six symptoms is too high for females and results in a high rate of false negatives. Therefore, our finding of a poor correspondence between symptom count cutoff and a scoring method based on normative data (dimensional approach) is consistent with the general literature.

In contrast to the findings for the disruptive behavior disorders, the categorical and dimensional methods of scoring yielded quite discrepant results for both Dysthymic Disorder and GAD. A possible explanation for this pattern is that *DSM-IV* indicates that specific symptoms must be present to meet criteria for GAD (anxiety, worry) or Dysthymic Disorder (depressed or irritable mood), whereas the disruptive behavior disorders only

require the presence of a specific number of symptoms, and all symptoms have equal importance. The absence of one or more of these *required* symptoms would affect the symptom count cutoff (categorical) but would not substantially impact the *T*-score cutoff (dimensional). One might also raise the issue of whether the different pattern is attributable to the nature of disruptive and affective disorders. Specifically, the symptoms of the former are much more observable to teachers than those of affective disorders. However, the fact that Social Phobia, one of the anxiety disorders, yielded quite good agreement between the two types of cutoffs challenges this explanation to some extent, although one could argue that symptoms of Social Phobia may be relatively more observable than those of Generalized Anxiety or Dysthymic Disorders.

DSM-IV is explicit in its requirement that without impairment, there is no disorder; therefore, the assessment of impairment is essential in the diagnostic process. One longstanding and relatively unchallenged inference of this requirement is that severity cutoff scores for behavior rating scales do not generate “true” estimates of disorder because severity is at best a modest indicator of impairment. Consistent with prior research (Gadow, Angelidou, et al., 2012; Rapee et al., 2012), we found the strongest agreement between ratings of symptom-induced impairment and symptom counts for disorders that are distressing to others, namely, ADHD-HI, ODD, and CD. In contrast, ADHD-I symptom ratings (especially of younger youth) were minimally associated with impairment ratings. This suggests that the presence of inattentive symptoms is only *part* of what determines impairment; that is, it is multiply determined and likely influenced by cognitive, environmental, and task variables. Furthermore, inattentive symptoms may be perceived as less impairing because their impact is limited to specific contexts such as academic settings. For yet other disorders, the severity of symptoms predicted impairment, but only when gender and/or age differences were taken into account. ADHD-HI is a case in point. For younger youth, the presence of six ADHD-HI symptoms among males or females predicted impairment. However, among the older cohort, only if a youth obtained a high *T*-score *relative to other same-gender peers* did it predict impairment beyond a fair degree. Given that the six-symptom cutoff for ADHD in the *DSM* is fixed, it is “blind” to gender- and age-related variations in symptom expression. Thus, *T*-score-based cutoffs are more associated with impairment when symptoms vary appreciably by age and/or gender. As to which of the two *T*-score cutoffs considered in our study showed better convergence with impairment ratings, $T \geq 70$ yielded higher kappas than $T \geq 65$ in most instances, which supports the continued use of $T \geq 70$ as a clinical cutoff.

In contrast, for other disorders, the *DSM* symptom count cutoffs were more associated with impairment ratings than *T*-score cutoffs. For Dysthymic Disorder, especially among younger youth, *T*-score cutoffs were only slightly related to impairment ratings (0.26–0.37), whereas the *DSM* symptom count cutoff (which required the presence of specific symptoms, notably depressed or irritable mood) had a more robust association (0.59–0.74) with impairment scores. This discrepancy may be due to the fact that aside from depressed or irritable mood, several of the remaining symptoms of Dysthymic Disorder are nonspecific to depressive mood and are more descriptive of inattention (e.g., trouble concentrating, low energy).

Strengths and Limitations

The strengths of this study include a large study sample of clinically referred youth, consideration of two different age groups, the inclusion of ratings of impairment, and the use of a single rating scale that was developed to be scored both categorically and dimensionally. Nevertheless, there are limitations as well. For example, the present study focused solely on teacher ratings of symptoms and impairment. The use of parent or youth self-reports would likely produce different findings. In the case of self-reports, the relationship between impairment and symptom ratings would be expected to be more similar for disruptive behavior disorders and affective disorders compared with that for parents' or teachers' ratings (Rapee et al., 2012). It should also be pointed out that this study addressed perceived symptom-induced impairment rather than independently assessed functional impairment. As Gordon et al. (2006) point out, at least for ADHD, there is only modest association between number or frequency of symptoms and various measures of functional impairment.

Clinical Implications

A common conundrum for clinicians is the need to resolve the seeming contradictions inherent in categorical and dimensional models of psychopathology as they are manifest in commonly used assessment instruments. This is no small matter as different scoring strategies based on these models identify different but overlapping groups of youth as meeting criteria for a specific condition, which has implications for eligibility for services. Interestingly, cutoff scores based on either approach yield similar findings most of the time in clinically referred youth, at least for the disruptive behavior disorders, but this is less true for the anxiety and depressive disorders examined in the present study, the explanation for which may be found in the multistep criteria (i.e., requiring specific symptoms) for the latter. Furthermore, the two methods of scoring yielded different patterns of association with impairment ratings, which suggests that each model provides unique information for clinical decision-making. These results are consistent with Mattison et al. (2003) who found that teachers' ratings of *DSM-IV* symptoms were more associated with their corresponding psychiatric diagnoses for the disruptive behavior disorders than for Dysthymic Disorder. For educators and health care professionals, these results support the use of validated assessment instruments based on both categorical and dimensional models.

Acknowledgments

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms and Profiles. Burlington: University of Vermont Research Center for Children, Youth, and Families; 2000.
- Altman, DG. Practical statistics for medical research. Boca Raton, FL: Chapman & Hall; 1997.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed., revised. Washington, DC: Author; 1987.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed.. Washington, DC: Author; 1994.

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., text rev. Washington, DC; Author: 2000.
- Bird HR, Davies M, Fisher P, Narrow WE, Jensen PS, Hoven C, Dulcan MK. How specific is specific impairment? *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39:1182–1189. [PubMed: 10986816]
- Boyle MH, Offord DR, Racine Y, Fleming JE, Szatmari P, Sanford M. Evaluation of the revised Ontario Child Health Study scales. *Journal of Child Psychology and Psychiatry*. 1993; 34:189–213. [PubMed: 8444992]
- Canino G, Shrout PE, Rubio-Stipec M, Bird HR, Bravo M, Ramirez R, Martinez-Taboas A. The DSM-IV rates of child and adolescent disorders in Puerto Rico. *Archives of General Psychiatry*. 2004; 61:85–93. [PubMed: 14706947]
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*. 1990; 43:551–558. [PubMed: 2189948]
- Coghill D, Sonuga-Barke EJS. Annual research review: Categories versus dimensions in the classification and conceptualization of child and adolescent mental disorders: Implications of recent empirical study. *Journal of Child Psychology and Psychiatry*. 2012; 53:469–489. [PubMed: 22288576]
- Conners, CK. Conners' Rating Scales-Revised. North Tonawanda, NY: Multi-Health Systems; 2002.
- Drabick AG, Gadow KD. Deconstructing oppositional defiant disorder: Clinic-based evidence for an anger/ irritability phenotype. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2012; 51:384–393. [PubMed: 22449644]
- Drabick AG, Gadow KD, Loney J. Source-specific oppositional defiant disorder: Comorbidity and risk factors in referred elementary schoolboys. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2007; 46:92–101. [PubMed: 17195734]
- Faraone SV, Sergeant J, Gillberg C, Biederman J. The worldwide prevalence of ADHD: Is it an American condition? *World Psychiatry*. 2003; 2:104–113. [PubMed: 16946911]
- Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*. 1990; 43:543–549. [PubMed: 2348207]
- Gadow KD, Angelidou K, Chernoff M, Williams PL, Heston J, Hodge J, Nachman S. Longitudinal study of emerging mental health concerns in youth perinatally-infected with HIV and peer comparisons. *Journal of Developmental & Behavioral Pediatrics*. 2012; 33:456–468. [PubMed: 22772819]
- Gadow KD, Chernoff M, Williams PL, Brouwers P, Morse E, Heston J, Nachman S. Co-occurring psychiatric symptoms in children perinatally infected with HIV and peer comparison sample. *Journal of Developmental & Behavioral Pediatrics*. 2010; 31:116–128. [PubMed: 20110828]
- Gadow KD, Drabick DAG. Anger and irritability symptoms among youth with ODD: Cross-informant versus source-exclusive syndromes. *Journal of Abnormal Child Psychology*. 2012; 40:1073–1085. [PubMed: 22581374]
- Gadow KD, Kaat AJ, Lecavalier L. Relation of symptom-induced impairment with other illness parameters in clinic-referred youth. Manuscript submitted for publication. 2012
- Gadow, KD.; Sprafkin, J. *Stony Brook Child Psychiatric Checklist*. Stony Brook, NY: Department of Psychiatry, Stony Brook University; 1986.
- Gadow, KD.; Sprafkin, J. *Child Symptom Inventories manual*. Stony Brook, NY: Checkmate Plus; 1994.
- Gadow, KD.; Sprafkin, J. *Adolescent supplement to the Child Symptom Inventories manual: Adolescent Symptom Inventory–4*. Stony Brook, NY: Checkmate Plus; 1995.
- Gadow, KD.; Sprafkin, J. *Child Symptom Inventory–4 screening and norms manual*. Stony Brook, NY: Checkmate Plus; 2002.
- Gadow, KD.; Sprafkin, J. *Child and Adolescent Symptom Inventory–4R: Teacher Checklist*. Stony Brook, NY: Checkmate Plus; 2005.
- Gadow, KD.; Sprafkin, J. *Adolescent Symptom Inventory–4 screening and norms manual*. Stony Brook, NY: Checkmate Plus; 2008.
- Gadow, KD.; Sprafkin, J. *The Symptom Inventories: An annotated bibliography*. Stony Brook, NY: Checkmate Plus; 2011. [On-line] Available from www.checkmateplus.com

- Gadow KD, Sprafkin J, Salisbury H, Schneider J, Loney J. Further validity evidence for the teacher version of the Child Symptom Inventory-4. *School Psychology Quarterly*. 2004; 19:50–71.
- Gathje RA, Lewandowski LJ, Gordon M. The role of impairment in the diagnosis of ADHD. *Journal of Attention Disorders*. 2008; 11:529–537. [PubMed: 18259000]
- Gordon M, Antshel K, Faraone S, Barkley R, Lewandowski L, Hudziak JJ, Cunningham C. Symptoms versus impairment: The case for respecting DSM-IV's Criterion D. *Journal of Attention Disorders*. 2006; 9:465–475. [PubMed: 16481663]
- Healey DM, Miller CJ, Castelli KL, Marks DJ, Halperin JM. The impact of impairment criteria on rates of ADHD diagnoses in preschoolers. *Journal of Abnormal Child Psychology*. 2008; 36:771–778. [PubMed: 18224434]
- Kazdin A, Kagan J. Models of dysfunction in developmental psychopathology. *Clinical Psychology: Science and Practice*. 1994; 1:35–52.
- Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: How reliable is reliable enough? *American Journal of Psychiatry*. 2012; 169:13–15. [PubMed: 22223009]
- Mattison RE, Gadow KD, Sprafkin J, Nolan EE, Schneider J. A DSM-IV-referenced teacher rating scale for use in clinical management. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2003; 42:442–449. [PubMed: 12649631]
- Orvaschel H, Puig-Antich J, Chambers W, Tabrizi MA, Johnson R. Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1982; 21:392–397.
- Pelham WE, Fabiano GA, Massetti GM. Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:449–476. [PubMed: 16026214]
- Pelham WE, Gnagy EM, Greenslade KE, Milich R. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1992; 31:210–218. [PubMed: 1564021]
- Ramtekhar UP, Reiersen AM, Todorov AA, Todd RD. Sex and age differences in attention-deficit/hyper-activity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2010; 49:217–228. [PubMed: 20410711]
- Rapee RM, Bogels SM, van der Sluis CM, Craske MG, Ollendick T. Annual research review: Conceptualising functional impairment in children and adolescents. *Journal of Child Psychology and Psychiatry*. 2012; 53:454–468. [PubMed: 22067073]
- Reich, W.; Shayka, JJ.; Taibleson, C. Diagnostic Interview for Child and Adolescents-Revised-Parent Version (DICA-R-P): DSM-III-R Version. St. Louis, MO: Washington University; 1991.
- Sprafkin J, Gadow KD, Salisbury H, Schneider J, Loney J. Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent Checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology*. 2002; 31:513–524. [PubMed: 12402570]
- Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Family Medicine*. 2005; 37:360–363. [PubMed: 15883903]
- Wakefield JC. When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Development and Psychopathology*. 1997; 9:269–290. [PubMed: 9201445]
- Winters NC, Collett BR, Myers KM. Ten-year review of rating scales, VII: Scales assessing functional impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2005; 44:309–338. [PubMed: 15782079]
- Wolraich ML, Feurer ID, Hannah JN, Baumgaertel A, Pinnock TY. Obtaining systematic teacher reports of disruptive behavior disorders utilizing DSM-IV. *Journal of Abnormal Child Psychology*. 1998; 26:141–152. [PubMed: 9634136]

Agreement (Kappa) Between Categorical (Symptom Count) and Dimensional (*T*-score) Scoring Algorithms for Children (Ages 6–12 Years) and Adolescents (Ages 13–18 Years)

Table 1

Disorder	Children (<i>n</i> = 516)						Adolescents (<i>n</i> = 576)					
	Males			Females			Males			Females		
	<i>T</i>	65	70	<i>T</i>	65	70	<i>T</i>	65	70	<i>T</i>	65	70
ADHD-I	.81		.56	.82		.93	.84		.47	.76		.92
ADHD-HI	.82		.82	.65		.69	.65		.85	.44		.50
ADHD-C ^a	.51		.76	.24		.33	.51		.74	.15		.21
ADHD-C ^b	.81		.62	.61		.71	.76		.71	.15		.23
ODD	.85		.93	.68		.73	.82		.93	.60		.68
CD	.78		.88	.43		.60	.87		.83	.34		.50
GAD	.42		.54	.37		.52	.18		.26	.13		.19
Social Phobia	.56		.56	.62		.62	— ^c		— ^c	— ^c		— ^c
Dysthymia	.32		.48	.22		.33	.45		.71	.40		.54

Note: ADHD-I = ADHD-Inattentive; ADHD-HI = ADHD-Hyperactive-Impulsive; ADHD-C = ADHD-Combined; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; GAD = Generalized Anxiety Disorder. All values significant at $p < .0001$.

^a ADHD-C *T*-score cutoff based on the total ADHD-C score meeting designated total *T*-score criterion.

^b ADHD-C *T*-score cutoff based on both ADHD-I and ADHD-HI scores meeting designated (*T* 65 or *T* 70) *T*-scores criterion.

^c The Social Phobia category contained too few items for an adequate normative distribution in adolescents.

Agreement (Kappa) of Impairment Cutoff Scores With Dimensional (*T*-Scores) and Categorical (*DSM-IV* Symptom Count) Cutoff Scores Among Child Psychiatric Referrals (Ages 6–12 Years)

Table 2

Disorder	Males (<i>n</i> = 372)				Females (<i>n</i> = 144)			
	<i>T</i>	<i>65</i>	<i>T</i>	<i>70</i>	<i>DSM-IV</i>	<i>T</i>	<i>65</i>	<i>DSM-IV</i>
ADHD-I	.22		.18		.26	.25 ^a	.22 ^a	.26 ^a
ADHD-HI	.65		.59		.73	.64	.65	.65
ADHD-C ^b	.49		.56		.61	.40	.48	.52
ODD	.81		.76		.76	.77	.82	.76
CD	.61		.73		.69	.53	.69	.82
GAD	.43		.53		.52	.45	.65	.64
Social Phobia	.58		.58		.48	.49	.49	.48
Dysthymia	.26		.32		.59	.26	.37	.74

Note: *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; APA, 1994); ADHD-I = ADHD-Inattentive; ADHD-HI = ADHD-Hyperactive-Impulsive; ADHD-C = Combined; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; GAD = Generalized Anxiety Disorder. All values significant at $p < .0001$ except where noted.

^a $p < .001$.

^b ADHD-C *T*-score cutoff based on the total ADHD-C score meeting designated total *T*-score criterion.

Table 3

Agreement (Kappa) of Impairment Cutoff Scores With Dimensional (*T*-Scores) and Categorical (*DSM-IV* Symptom Count) Cutoff Scores Among Adolescent Psychiatric Referrals (Ages 13–18 Years)

Disorder	Males (<i>n</i> = 391)				Females (<i>n</i> = 185)					
	<i>T</i>	65	<i>T</i>	70	<i>DSM-IV</i>	<i>T</i>	65	<i>T</i>	70	<i>DSM-IV</i>
ADHD-I		.32		.37		.57		.25	.32	.69
ADHD-HI		.72		.36		.32		.65	.72	.26
ADHD-C ^a		.51		.38		.28		.38	.47	.16
ODD		.82		.79		.76		.75	.75	.60
CD		.64		.50		.60		.47	.55	.26
GAD		.44		.55		.41		.42	.52	.37
Dysthymia		.33		.54		.60		.46	.54	.56
Social Phobia		__ <i>b</i>		__ <i>b</i>		.33		__ <i>b</i>	__ <i>b</i>	.41

Note: *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; APA, 1994); ADHD-I = ADHD-Inattentive; ADHD-HI = ADHD-Hyperactive-Impulsive; ADHD-C = ADHD-Combined; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; GAD = Generalized Anxiety Disorder. All values significant at $p < .0001$.

^a ADHD-C *T*-score cutoff based on the total ADHD-C score meeting designated total *T*-score criterion.

^b The Social Phobia category contained too few items for an adequate normative distribution in adolescents.