

The Relation Between Change in Symptoms and Functioning in Children with ADHD Receiving School-Based Mental Health Services

Julie Sarno Owens · Lindsay M. Johannes ·
Veronika Karpenko

Published online: 19 September 2009
© Springer Science+Business Media, LLC 2009

Abstract This study examined the relation between reliable change in symptoms and reliable change in functioning in children with attention deficit hyperactivity disorder (ADHD; $N = 64$) who were enrolled in a school-based mental health program that included a daily report card intervention, teacher consultation, and behavioral parenting sessions. Parents and teachers completed the disruptive behavior disorders rating scale and the impairment rating scale at pre- and post-treatment. Group-level analyses indicated that symptom improvers demonstrated significant improvement in multiple domains of functioning, whereas symptom no-changers and deteriorators did not. However, individual-level analyses revealed that up to 40% (depending on the domain and the informant) of children achieved reliable symptom change without reliable improvement in functioning, and up to 16% achieved reliable change in functioning without reliable change in symptoms. The results have implications for measurement of treatment outcome.

Keywords ADHD · Clinical significance · Reliable change index · Treatment outcome · Functioning, school-based mental health

Children with attention deficit hyperactivity disorder (ADHD) demonstrate developmentally inappropriate levels of inattention, and/or hyperactivity and impulsivity that often interfere with their ability to perform well academically, and to initiate and maintain positive relationships

with peers, siblings, and adults (Barkley, 2006). Recent reviews (Pelham & Fabiano, 2008; Pelham, Wheeler, & Chronis, 1998) and meta-analyses (Fabiano et al., 2009) offer strong and consistent evidence for the effectiveness of behavioral treatments for children with ADHD. These articles provide a comprehensive evaluation of extant studies, with indisputable value; however, they have focused overwhelmingly on statistically significant change, effect size magnitude, and group-level data. Very few of the reviews have provided information about the reliability or clinical meaningfulness of the change that occurs at the individual level. Additionally, although recent studies have started to assess a variety of outcome measures (e.g., symptoms, functioning, satisfaction), very few have analyzed the correspondence between different outcome measures (Karpenko, Owens, Evangelista, & Dodds, 2009).

This problem is not specific to ADHD treatment outcome studies. Indeed, most psychotherapy research has relied on the assessment of symptoms to measure treatment outcome and has utilized inferential statistical analyses to draw conclusions about an average client (Ogles, Lunnen, & Bonesteel, 2001). A more clinically relevant indicator of treatment outcome is *clinically significant (CS) change*, defined as change in treatment that is meaningful and noticeable to the *individual* client or to significant people in the client's life (Jacobson & Truax, 1991; Kazdin, 1999). Jacobson and Truax's (1991) formula is one of the most frequently used methodologies for defining CS change. Using this formula, researchers can determine whether an individual made (a) reliable change from pre- to post-treatment and (b) whether the change made by the client places him/her in the normative distribution on a given dimension (e.g., symptom severity). Using both criteria, a client is classified as making CS change if he/she made reliable change *and* has post-treatment scores in the normative range.

J. S. Owens (✉) · L. M. Johannes · V. Karpenko
Department of Psychology, Ohio University, Porter Hall 200,
Athens, OH 45701, USA
e-mail: owensj@ohio.edu

However, the two-criterion definition of CS change has been criticized as being overly conservative (e.g., Tingey, Lambert, Burlingame, & Hansen, 1996); as such, many studies measuring clinical significance use the reliable change criterion only (Ogles et al., 2001; Rosenblatt & Rosenblatt, 2002). The reliable change index (RCI) classifies clients into three categories based on the direction and the magnitude of change (improvers, no-changers, and deteriorators) regardless of return to the normal range of functioning. Alternative computational approaches for calculating RCI have been proposed; however, the broad conclusion is that these alternatives have more similarities than differences (see Wise, 2004 for a review) that none offer a significant benefit over the Jacobson and Truax method, and for consistency, researchers should continue to use the Jacobson and Truax method. Further, because most empirically-supported treatments for ADHD result in significant improvement, but not normalized functioning for many children (e.g., Swanson et al., 2001), the most conservative examination of CS change may have less utility than an examination that is applicable to a broader range of children. For these reasons, Jacobson and Truax's RCI method was used in this study.

To date, most studies have examined CS change in *symptoms*, and very few studies have examined the relation between change in symptoms and other important outcome measures (e.g., domains of functioning). This distinction is noteworthy given the modest association between symptoms and impairment (Gordon et al., 2006). An examination of treatment-related change in functioning and the connection between change in symptoms and change in functioning is particularly important in the treatment of ADHD when one considers the chronic nature of the disorder, the multiple domains of functioning that are impaired, and the diagnostic criteria.

The diagnostic criteria for ADHD (APA, 2000, p. 92) necessitate the assessment of functional impairment in multiple settings and from multiple perspectives (Evans & Youngstrom, 2006; Gordon et al., 2006). For children with ADHD, impairment often occurs at school (e.g., academic impairment, organizational difficulties, disruptions in peer relations, strained teacher–child relations) and at home (e.g., inability to complete household chores, excessive injuries due to impulsivity, and strained relations with parents). In fact, it is these impairments, rather than the symptoms of ADHD, that often lead adults to refer children to treatment (Angold, Costello, Farmer, Burns, & Erkanli, 1999; Fabiano et al., 2006). Furthermore, indicators of functioning such as academic performance and peer relationships are particularly strong predictors of long-term adjustment (e.g., Barkley, Fischer, Smallish, & Fletcher, 2006; Parker & Asher, 1987). Thus, given the ecological relevance of indicators of functioning, the importance of

impairment in the diagnostic determination of ADHD, and the breadth of impaired domains for children with the disorder, greater attention must be focused on the impact of treatment on functioning, as well as the correspondence between change in symptoms and change in functioning.

To our knowledge, there are only two studies that have examined this correspondence (Karpenko et al., 2009; Rosenblatt & Rosenblatt, 2002). Rosenblatt and Rosenblatt (2002) examined the agreement between reliable change in youth- and parent-reported symptoms on the child behavior checklist and youth self report (Achenbach, 1991a, b) and therapist-rated functioning on the child and adolescent functional assessment scale (Hodges, 1994) in adolescents receiving services at a community mental health clinic. Results indicated minimal correspondence between change in symptoms and functioning, with a sizable minority of youth (33%) making reliable change in therapist-rated functioning without reliable change in youth-rated symptoms. This study has two noteworthy limitations. Within-informant analyses (i.e. the correspondence of symptoms and functioning for each informant) were not conducted and the study measured global functioning, rather than functioning within specific domains, possibly limiting the implications for treatment modification or planning.

Karpenko et al. (2009) improved upon these limitations. Using data from the multimodal treatment study of children with ADHD (MTA; MTA Cooperative Group, 1999), this study examined the relation between clinically significant change in symptoms and reliable change in five domains functioning, five of which were rated by parents and only one of which was rated by teachers (social skills). Although there was statistically significant correspondence between change in symptoms and change in functioning, up to 52% of children (depending on the domain of functioning) who did not achieve CS change in symptoms showed reliable improvement in functioning indicators. These findings raise important implications for the definition of successful treatment outcome, as reliable change in an impaired domain of functioning may be meaningful to clients even in the absence of significant improvement in symptoms. One of the limitations of the Karpenko et al. (2009) study is that only one area of functioning was rated by the teachers, despite that children with ADHD demonstrate multiple functional impairments in the school setting.

The present study examined the correspondence between reliable change in symptoms and functioning in children with ADHD receiving school-based mental health services, extending the current literature in several ways. First, the present study incorporates both parent and teacher report of functioning across multiple domains, utilizing parallel forms of the same measure of functioning. Second, it examines the correspondence between symptoms and functioning using both group- and individual- level

analyses. The results of these analyses highlight the potentially inaccurate conclusions that could be drawn about such correspondence if results were based solely on group-level analyses. Finally, the present study examines the correspondence between symptoms and functioning using both within-informant and cross-informant analyses. Neither Karpenko et al.'s (2009) study nor Rosenblatt and Rosenblatt's (2002) study included *both* within-informant and cross-informant analyses at a group and individual level. It is important to analyze both within- and cross-informant findings to both to ensure that any associations found are not inflated due to within-source variance.

Methods

Participants

Participants in the current analyses were 64 children in kindergarten through 6th grade who were enrolled in The Youth Experiencing Success in School (Y.E.S.S.) Program across 6 years (see Table 1 for demographic characteristics). This collaborative school mental health program (Owens, Murphy, Richerson, Girio, & Himawan, 2008; Owens et al., 2005) provides empirically-supported treatments for ADHD and oppositional defiant disorder (ODD), including a daily report card procedure (Kelley, 1990),

year-long behavioral teacher consultation (e.g., Sheridan, Kratochwill, & Bergan, 1996), and behavioral parenting sessions (Barkley, 1997). These interventions were available across the entire academic year (see Owens et al., 2008 for details). Across children, the daily report card was implemented, on average, for 67 school days ($SD = 35.51$) with teachers implementing the DRC procedures on 75% of eligible school days ($SD = 19.17$). On average, parents received 10 h of parent–clinician contact ($SD = 5.36$), and teachers received 8.12 h of teacher–clinician consultation ($SD = 4.23$). This study presents data on new enrollees only (i.e., no case represents a child who repeated the program).

The school districts in which the study took place are located in low-income communities within the Appalachian region of Ohio where county statistics indicate that the child poverty rate, the unemployment rate, the uninsured rate, and the percentage of students who qualify for free and reduced lunches (50%, on average, across buildings) exceed state rates.

Diagnostic status was determined using parent and teacher versions of the disruptive behavior disorders (DBD) rating scale (Pelham, Gnagy, Greenslade, & Milich, 1992) and the impairment rating scale (IRS; Fabiano et al., 2006), in combination with a semi-structured parent interview conducted by a graduate student clinician (either the disruptive behavior disorders structured parent interview (Pelham, 2002) or children's interview for psychiatric syndromes- parent version (Weller, Weller, Teare, & Fristad, 1999) depending on the year). Given the strength of their psychometric properties, priority was given to data from the parent and teacher rating scales (Pelham, Fabiano, & Massetti, 2005) when making diagnostic decisions. To meet criteria for ADHD, six or more symptoms of either inattention or hyperactivity/impulsivity had to be endorsed (as "pretty much" present or "very much" present) on the DBD rating scales. The symptoms may have been endorsed by the teacher, the parent, or a combination. The same symptom was not counted twice if endorsed by both raters. In addition, both the parent and the teacher had to endorse impairment (a score of 3 or higher on the IRS) in at least one domain. To meet the criteria for oppositional defiant disorder (ODD), four or more symptoms of ODD had to be endorsed on the DBD rating scale by either parent or teacher. To meet the criteria for conduct disorder (CD), 3 or more symptoms had to be endorsed. For the diagnoses of ODD and CD, impairment had to be endorsed by one rater. In the few cases in which the parent interview resulted in conflicting information, the program clinician and the licensed clinical supervisor resolved these discrepancies by incorporating other data (e.g., teacher interview, behavioral observation) and contextual information (e.g., possible rater biases).

Table 1 Demographic characteristics of participants (total $N = 64$)

| Variable | N (%) |
|------------------------------|---------------|
| Gender (% male) | 56 (87.5) |
| Race | |
| White | 55 (85.9) |
| Black | 1 (1.6) |
| Native American | 3 (4.7) |
| Other/biracial | 2 (3.1) |
| Grade | |
| K through 3rd grade | 47 (73.4) |
| 4th, 5th, or 6th grade | 17 (26.6) |
| On medication at referral | 21 (32.8) |
| In counseling at referral | 21 (32.8) |
| Receiving special education | 35 (54.7%) |
| Repeated a grade | 13 (20.3) |
| Medication part of treatment | 32 (50.0) |
| IQ estimate (M , SD) | 95.02 (13.66) |
| SES (M , SD) | 26.09 (12.68) |

Symptom improvers and symptom no-changers (based on parent ratings and teacher ratings) do not differ on medication status at referral or during treatment. The IQ estimate was derived from the Weschler abbreviated scales of intelligence (WASI) 2-subtest scale
SES Socioeconomic status

Of the 64 children in the sample, 13 children (20%) met criteria for ADHD alone, and 51 children (80%) met criteria for ADHD and a co-occurring behavioral or mood disorder. According to parent-report (demographic questionnaire), 33% were taking medication at the time of referral. According to clinician report at the end of the year, 50% of children had received medication as part of their treatment during the school year. However, compliance with the recommended medication dose and schedule were not monitored.

Complete pre-treatment (Fall of the school year) and post-treatment (Spring of the school year) teacher ratings of symptoms and functioning were available for 57 of the 64 children. All children included in the teacher analyses had Time 1 teacher ratings of ADHD symptoms that were rated as 1 (“just a little”) or higher, indicating symptomatology that was outside of the normative range. This criterion was selected to remain consistent with other studies (Karpenko et al., 2009; Swanson et al., 2001). Pre- and post-treatment parent ratings of symptoms and functioning were available for 39 children.¹ All children included in the parent analyses had Time 1 parent ratings of ADHD symptoms that were rated as 1 (“just a little”) or higher. Thus, the within-informant analyses include 39 cases for parent analyses and 57 for teacher analyses. The cross-informant analyses include 43 cases when using parent-based symptoms groups and 33 cases when using teacher-based symptom groups.

¹ Of the 25 children for whom there are teacher data but not parent data, 19 (76%) are due to lack of parent completion of one of the rating scales (as ratings of symptoms and impairment at both time points are needed to remain in the parent analyses). This was primarily due to the fact that parents were not paid for rating scale completion in the first year of the program, resulting in more missing data in year 1 (for 7 of 19 cases) than in other years (12 spread across the remaining 5 years). The remaining 6 children (24%) were not included in the parent analyses because parent ratings of symptoms fell within the normal range. Early identification and early intervention are priorities of the school mental health program (e.g. 73% of the sample is in K through 3rd grade). As such, it is common for teacher ratings to be more severe than parent ratings at baseline, as this is often the first time that parents have been informed that their child is impaired or having difficulty. Similarly, in our clinical experience with this low-income rural sample, it is common for parent report of child symptoms to be slightly less severe on paper rating scales than during the face-to-face diagnostic interview (perhaps due to greater comfort or less defensiveness in a face-to-face meeting). Because multiple sources of information (i.e., interviews, rating scales of symptoms and impairment, and observation of the child) are synthesized to determine the child’s diagnosis, these children meet the criteria for ADHD. However, to remain consistent with the criteria for the within-informant teacher analyses, these children are not included in the within-informant parent analyses due to the parent rating of symptoms at baseline. Thus, the sample sizes vary analyses by analyses.

Independent samples *t*-test analyses indicated that children with complete parent data did not differ significantly from those without complete parent data on socioeconomic status (Hollingshead, 1975), cognitive ability tests scores, or teacher-rated ADHD or ODD symptoms.

Measures

Disruptive Behavior Disorders (DBD) Rating Scale

The DBD rating scale (Pelham et al., 1992) is a 45-item measure that assesses DSM-IV-based symptoms of inattention, hyperactivity/impulsivity, oppositional defiant disorder, and conduct disorder. Parents and teachers rate the severity of each symptom on a 4-point scale ranging from 0 (“not at all” present) to 3 (“very much” present). This rating scale has strong psychometric properties including high internal consistency of each factor, respectable test–retest reliability, and strong evidence of convergent validity and sensitivity to change (Pelham, Fabiano, & Massetti, 2005). For this study, an ADHD subscale was created by averaging the inattention and hyperactivity subscales for parent and teachers separately (alphas were .92 and .90, respectively).

Impairment Rating Scale (IRS)

The IRS (Fabiano et al., 2006) assesses adult perceptions of child functioning in multiple domains (academic performance, classroom functioning, family functioning, self-esteem, and relationships with peers, siblings, parents, and teachers). Parents and teachers rate the severity of the child’s impairment in each domain on a 7-point scale, ranging from 0 (no problem) to 6 (extreme problem). The measure has respectable cross-informant reliability (e.g., correlations above .47), convergent and divergent validity with other impairment scales (e.g., correlation of .77 between IRS overall impairment and the children’s global assessment scale), and predictive validity in identifying children with ADHD diagnoses (Fabiano et al., 2006). Test retest reliabilities of the six items range from $r = .75$ to $.90$ on the parent version, and from $r = .65$ to $.91$ on the teacher version (Fabiano et al., 2006). The item assessing sibling relations is not included in the current analyses due to smaller sample size.

Procedure

The data analyzed represent parent and teacher DBD and IRS scores from pre-treatment (Time1) and post-treatment (Time2). Using Jacobson and Truax’s (1991) methodology, a reliable change index (RCI) was created for the ADHD symptoms subscale for each informant, as well as for each

IRS domain for each informant. The following formula was used:

$$RCI = X_{\text{post}} - X_{\text{pre}}/S_{\text{diff}}; S_{\text{diff}} = \sqrt{2(S_E)^2};$$

$$S_E = SD_1\sqrt{1 - r_{xx}}.$$

In this formula, r_{xx} = test–retest reliability of the measure; S_{diff} = standard error of the difference between the two test scores; SD_1 = standard deviation of the present sample at time one; S_E = standard error of measurement. If the RCI is greater or equal to 1.96, children are considered deteriorators, if the RCI is less than or equal to -1.96 , children are categorized as improvers, and if the RCI falls between 1.96 and -1.96 , children are considered no-changers. Based on parent-rated ADHD symptoms, 39% of children ($N = 15$) were categorized as improvers, 41% ($N = 16$) as no-changers, and 21% ($N = 8$) as deteriorators. Based on teacher-rated ADHD symptoms, 47% of children ($N = 27$) were categorized as improvers, 44% ($N = 25$) as no-changers, and 9% ($N = 5$) as deteriorators. Due to the small number of deteriorators, deteriorators and no-changers were combined into one group (henceforth referred to as “symptom no-changers” or “functioning no-changers”).

Independent sample t -test results indicated that parent-rated symptom improvers had more severe Time 1 parent-rated ADHD symptoms ($M = 2.02$; $SD = 0.44$) than did symptom no-changers ($M = 1.65$; $SD = 0.46$), $t(37) = -2.46$, $p < .05$, and more parent–clinician contact hours ($M = 13.87$; $SD = 6.85$) than did symptom no-changers ($M = 8.65$; $SD = 3.67$), $t(33) = -2.95$, $p < .01$. Similarly, the duration of the DRC intervention (in school days) was marginally longer for the parent-rated symptom changers ($M = 83.33$; $SD = 37.42$) than for symptom no-changers ($M = 58.17$; $SD = 35.37$), $p < .08$. There were no significant group differences on any parent-rated domain of functional impairment at Time 1, on the teacher compliance with the DRC intervention (i.e. percentage of school days implemented as intended), or on the number of teacher–clinician contact hours.

T -tests conducted on teacher-rated symptom improvement groups indicated that there were no significant group differences on Time 1 teacher-rated ADHD symptom severity, on any domain of impairment, or on any intervention dose variable including, teacher compliance with the DRC intervention, DRC duration, number of parent–clinician contact hours, and number of teacher–clinician contact hours.

Results

Two sets of analyses were conducted. The first set examines the correspondence between symptoms and functioning at the *group* level, first within informant, and then

across informants. The second set examines the correspondence between symptoms and functioning at the *individual* child level, first within informant, and then across informant.

Correspondence at the Group Level

Within-Informant Analyses

A 2 (time: pre-treatment, post-treatment) \times 2 (group: symptom improvers, symptom no-changers) repeated measures MANOVA examined differences between the two symptom groups on the six functional domains of the IRS over time. The analyses were conducted on parent and teacher reports separately.

For parent-report, significant multivariate effects were found for time, Wilks' Lambda = .76, $F(1, 31) = 10.01$, $p < .005$; for IRS domain, Wilks' Lambda = .39, $F(5, 27) = 8.60$, $p < .001$; and for the Group \times Time interaction, Wilks' Lambda = .79, $F(1, 31) = 8.35$, $p < .05$. Simple effects for the two-way interaction indicated that the groups do not differ at Time 1, but do differ at Time 2 [$F(1, 31) = 4.33$, $p < .05$]. In addition, collapsed across IRS domains, the symptom improvers evidenced a significant reduction in impairment across time [$F(1, 31) = 14.40$, $p < .001$], whereas the symptom no-changers did not change significantly over time (see Fig. 1). Means and standard deviations for parent data are presented in Table 2 (see parent columns). The significant main effect of domain revealed that, collapsed across time, the IRS overall impairment item was rated as significantly more severe than all of the specific domain items, with the exception of academic impairment. Academic impairment was rated as significantly more severe than peer impairment.

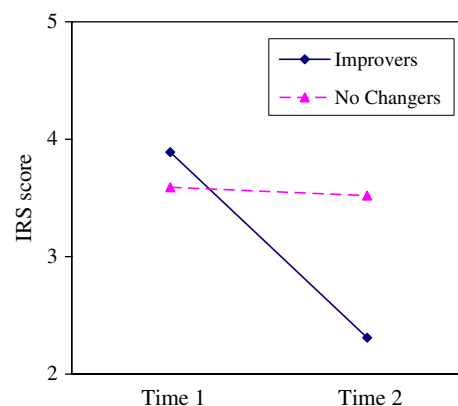


Fig. 1 Changes in functional impairment collapsed across domains by symptom-based reliable change group—within informant parent-report

Table 2 Descriptive statistics for parent-rated symptom groups across time

| | Parent (<i>N</i> = 39) | | Teacher (<i>N</i> = 43) | |
|----------------------------------|---------------------------------|----------------------------------|---------------------------------|----------------------------------|
| | Pre (<i>M</i> , <i>SD</i>) | Post (<i>M</i> , <i>SD</i>) | Pre (<i>M</i> , <i>SD</i>) | Post (<i>M</i> , <i>SD</i>) |
| Improvers | | | | |
| ADHD symptoms | 2.02 (0.44) | 0.98 (0.51) | 1.88 (0.73) | 1.36 (0.64) |
| IRS peers | 3.50 (1.68) | 1.75 (1.42) | 4.11 (1.45) | 3.16 (2.09) |
| IRS parent/teacher | 4.08 (2.02) | 2.08 (2.15) | 3.84 (1.71) | 3.05 (2.01) |
| IRS academics | 4.25 (2.09) | 2.67 (2.27) | 5.21 (1.27) | 3.89 (1.94) |
| IRS family/classroom | 3.83 (1.75) | 2.17 (2.25) | 4.68 (1.38) | 3.16 (2.12) |
| IRS self-esteem | 3.08 (2.15) | 2.67 (2.10) | 4.47 (1.58) | 3.74 (1.94) |
| IRS overall | 4.58 (1.24) | 2.50 (2.28) | 5.00 (1.16) | 3.53 (1.90) |
| No-changers/deteriorators | | | | |
| ADHD symptoms | 1.65 (0.46) | 1.86 (0.58) | 2.05 (0.64) | 1.95 (0.59) |
| IRS peers | 3.14 (1.39) | 3.05 (1.63) | 4.71 (1.00) | 4.29 (1.65) |
| IRS parent/teacher | 3.62 (2.25) | 3.19 (1.81) | 4.21 (1.87) | 3.88 (1.90) |
| IRS academics | 3.52 (2.09) | 3.81 (1.99) | 4.96 (1.33) | 5.04 (1.58) |
| IRS family/classroom | 3.38 (1.99) | 3.19 (2.06) | 4.96 (0.86) | 4.71 (1.30) |
| IRS self-esteem | 3.71 (2.03) | 3.71 (1.98) | 4.29 (1.40) | 4.29 (1.37) |
| IRS overall | 4.14 (1.28) | 4.14 (1.68) | 5.08 (0.83) | 4.79 (1.47) |

Parent columns represent within informant data. Teacher columns represent cross-informant data

IRS Impairment rating scale

For teacher-report, significant multivariate effects were found for time, Wilks' Lambda = .65, $F(1, 54) = 29.58$, $p < .01$; for IRS domain, Wilks' Lambda = .42, $F(5, 50) = 13.91$, $p < .001$; and for the Group \times Time interaction, Wilks' Lambda = .57, $F(1, 54) = 41.51$, $p < .001$. Similar to the parent data, the simple effects test for the Group \times Time interaction revealed that the groups do not differ at Time 1, but do differ at Time 2 ($F(1, 54) = 23.73$, $p < .05$). In addition, collapsed across IRS domains, the symptom improvers evidenced a significant reduction in impairment across time, [$F(1, 54) = 65.88$, $p < .001$], whereas the symptom no-changers did not change significantly over time (see Fig. 2). Means and standard deviations for teacher data are presented in Table 3 (see teacher columns). The significant main effect of domain revealed that, collapsed across time, the IRS overall impairment item, academic impairment, and classroom functioning items were rated as significantly more severe than all other items.

Cross-Informant Analyses

The first cross-informant analysis was a 2 (time) \times 2 (group) repeated measures MANOVA where group was represented by *teacher-rated symptom groups* and the dependent variables were *parent-rated functioning scores*. Significant multivariate effects were found only for time, Wilks' Lambda = .76, $F(1, 31) = 9.78$, $p < .01$; and for

IRS domain, Wilks' Lambda = .34, $F(5, 27) = 10.36$, $p < .001$. These effects revealed that Time 1 ratings were more severe than Time 2 ratings for both groups (see Fig. 3) and that parent ratings of academic impairment and overall impairment were more severe than ratings of impairment with peers, parents, self-esteem and family functioning. Means and standard deviations for these analyses are presented in Table 3 (see parent columns).

The second cross-informant analysis was a 2 (Time) \times 2 (Group) repeated measures MANOVA where group was represented by *parent-rated symptom groups* and the

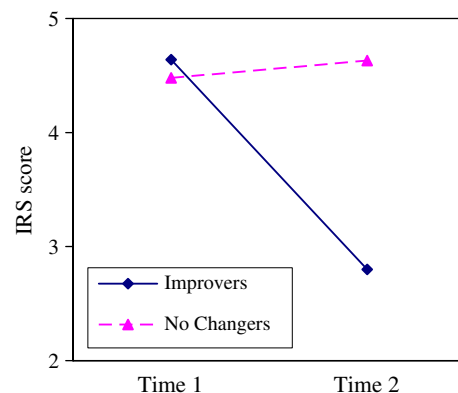


Fig. 2 Changes in functional impairment collapsed across domains by symptom-based reliable change group—within informant teacher-report

Table 3 Descriptive statistics for teacher-rated symptom groups across time

| | Parent (N = 33) | | Teacher (N = 57) | |
|----------------------------------|-----------------|--------------|------------------|--------------|
| | Pre (M, SD) | Post (M, SD) | Pre (M, SD) | Post (M, SD) |
| Improvers | | | | |
| ADHD symptoms | 1.72 (0.73) | 0.90 (0.51) | 2.21 (0.46) | 1.20 (0.53) |
| IRS peers | 2.83 (2.04) | 1.58 (1.56) | 4.54 (1.14) | 2.46 (1.88) |
| IRS parent/teacher | 3.33 (1.97) | 1.67 (1.92) | 4.12 (1.63) | 2.27 (1.89) |
| IRS academics | 3.50 (2.02) | 2.25 (2.34) | 5.19 (1.02) | 3.58 (2.19) |
| IRS family/classroom | 3.50 (1.57) | 2.33 (2.27) | 5.04 (.92) | 2.73 (1.87) |
| IRS self-esteem | 2.50 (2.11) | 2.17 (2.25) | 3.77 (1.97) | 2.85 (2.05) |
| IRS overall | 4.08 (1.38) | 2.33 (2.23) | 5.15 (0.73) | 2.88 (2.03) |
| No-changers/deteriorators | | | | |
| ADHD symptoms | 1.54 (0.61) | 1.61 (0.66) | 2.02 (0.50) | 2.04 (0.53) |
| IRS peers | 2.71 (1.65) | 2.33 (1.71) | 3.90 (1.95) | 4.40 (1.52) |
| IRS parent/teacher | 3.38 (2.50) | 2.52 (2.09) | 3.87 (1.93) | 4.10 (1.61) |
| IRS academics | 4.19 (1.89) | 4.24 (1.38) | 5.23 (1.14) | 5.37 (.96) |
| IRS family/classroom | 3.14 (2.20) | 2.43 (2.18) | 4.40 (1.75) | 4.50 (1.59) |
| IRS self-esteem | 3.62 (2.09) | 3.38 (2.01) | 4.37 (1.45) | 4.30 (1.39) |
| IRS overall | 3.86 (1.65) | 3.86 (1.71) | 5.10 (.89) | 5.13 (.97) |

Teacher columns represented within informant data
 IRS impairment rating scale

dependent variables were *teacher-rated functioning scores*. Significant multivariate effects were found for time, Wilks' Lambda = .78, $F(1, 41) = 11.59, p < .01$; and for IRS domain, Wilks' Lambda = .51, $F(5, 37) = 7.09, p < .001$, and for the Group \times Time interaction, Wilks' Lambda = .88, $F(1, 41) = 5.65, p < .05$. The simple effects test for the Group \times Time interaction revealed that the groups do not differ at Time 1, but do differ at Time 2 [$F(1, 41) = 5.35, p < .05$]. In addition, collapsed across IRS domains, the symptom improvers evidenced a significant reduction in impairment across time, [$F(1, 41) = 14.97,$

$p < .001$], whereas the symptom no-changers did not change significantly over time (see Fig. 4). Means and standard deviations for teacher data are presented in Table 2 (see teacher columns).

Correspondence at the Individual Level

A series of Chi-square and McNemar tests were conducted on the symptom groups (symptom improvers, symptom no-changers) and the functioning groups (functioning improvers, functioning no-changers) for each domain.

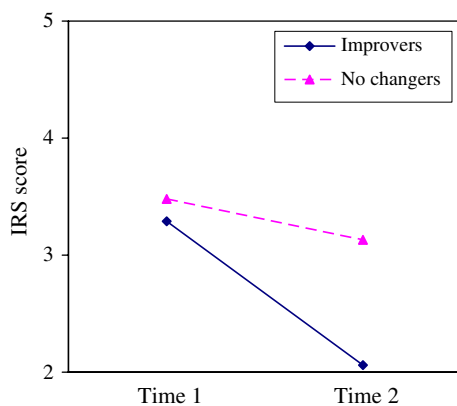


Fig. 3 Changes in parent-rated functional impairment collapsed across domains by teacher-rated symptom-based reliable change group

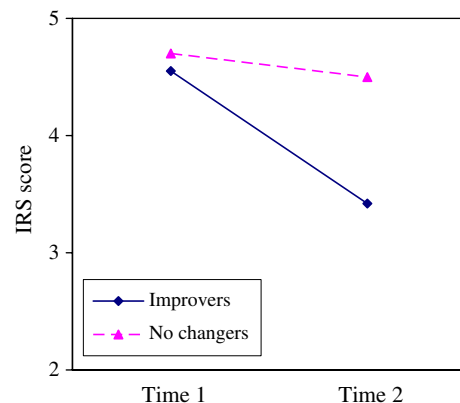


Fig. 4 Changes in teacher-rated functional impairment collapsed across domains by parent-rated symptom-based reliable change group

Table 4 Correspondence between symptom and functioning groups within and across informant

| Domain | Parent ratings | | Teacher ratings | |
|--|-------------------------|----------------------|-------------------------|----------------------|
| | Symptoms No-changers | Symptom Improvers | Symptoms No-changers | Symptom Improvers |
| Parent IRS | | | | |
| Peer no-changers ^a | 61 | 25 ^c | 61 | 27 ^c |
| Peer improvers | 3 (4) | 11 (31) | 3 (5) | 9 (25) |
| Parent–child no-changers | 47 | 21 | 49 | 21 |
| Parent–child improvers | 16 (25) | 16 (43) | 15 (24) | 15 (42) |
| Academic no-changers | 56 | 28 ^c | 55 | 27 |
| Academic improvers | 6 (9) | 11 (29) | 9 (14) | 9 (25) |
| Self-esteem no-changers | 58 | 27 | 61 | 30 ^c |
| Self-esteem improvers | 6 (10) | 9 (25) | 3 (5) | 6 (17) |
| Family no-changers ^a | 56 | 25 | 55 | 27 |
| Family improvers | 6 (9) | 14 (36) | 9 (14) | 9 (25) |
| Overall no-changers ^a | 53 | 18 | 52 | 18 |
| Overall improvers | 8 (13) | 21 (53) | 12 (19) | 18 (50) |
| Teacher IRS | | | | |
| Peer no-changers ^a | 51 | 33 ^c | 54 | 27 ^c |
| Peer improvers | 5 (8) | 12 (26) | 0 (0) | 20 (42) |
| Teacher–child no-changers ^a | 51 | 40 ^c | 54 | 34 ^c |
| Teacher–child improvers | 5 (8) | 5 (11) | 0 (0) | 13 (27) |
| Academic no-changers ^a | 51 | 30 ^c | 50 | 27 |
| Academic improvers | 5 (8) | 14 (32) | 4 (7) | 20 (43) |
| Self-esteem no-changers ^{a,b} | 56 | 37 ^c | 54 | 36 ^c |
| Self-esteem improvers | 0 (0) | 7 (16) | 0 (0) | 11 (23) |
| Classroom no-changers ^a | 47 | 26 | 46 | 18 |
| Classroom improvers | 9 (17) | 19 (42) | 7 (13) | 29 (62) |
| Overall no-changers ^{a,b} | 44 | 19 | 42 | 12 |
| Overall improvers | 12 (21) | 26 (58) | 11 (20) | 35 (74) |

Values are given in percentages (based on the total sample). Those in parentheses represent the percentage *within each symptom group* who made reliable change in functioning

^a $p < .05$ for within-informant Chi-square analyses

^b $p < .05$ for cross-informant Chi-square analyses

^c $p < .05$ for McNemar analyses

Within-Informant Analyses

Results for within informant parent-report suggest that there was a significant association between parent-rated improvement in symptoms and parent-rated improvement in functioning for three of six domains: peer relations, $\chi^2(1, N = 36) = 4.85, p < .03$; Family functioning, $\chi^2(1, N = 36) = 3.87, p < .05$; and overall, $\chi^2(1, N = 38) = 7.17, p < .01$ (see Table 4 parent column). Despite these significant associations, there were a number of children for whom change in symptoms did not correspond with change in functioning (see off diagonal in Table 4, parent column). Thus, follow-up McNemar tests were conducted to examine which type of discordance was more

likely (i.e. change in symptoms without change in functioning or change in functioning without change in symptoms). Results were significant for the peer relations ($p < .03$) and academic domains ($p < .04$), indicating that there were significantly more youth who made reliable change in parent-rated symptoms without reliable change in these domains for functioning (25 and 28%, respectively) than youth who made reliable change in functioning without reliable change in symptoms (3 and 6%, respectively).

The Chi-square results for the within-informant teacher report revealed a significant association between symptoms and functioning for all six domains: peer relations, $\chi^2(1, N = 56) = 15.80, p < .001$; teacher–child relations, $\chi^2(1, N = 56) = 9.23, p < .01$; academic, $\chi^2(1, N = 56) =$

9.93, $p < .01$; classroom functioning, $\chi^2(1, N = 46) = 14.10$, $p < .001$; self-esteem, $\chi^2(1, N = 56) = 7.75$, $p < .01$; overall, $\chi^2(1, N = 57) = 16.75$, $p < .001$ (see Table 4 teacher column).

McNemar tests were significant for the peer relations ($p < .001$), teacher–child relations ($p < .001$), academic ($p < .01$), and self-esteem domains ($p < .001$), indicating that there were significantly more youth who made reliable change in symptoms without reliable change in functioning than youth who made reliable change in functioning without reliable change in symptoms (see Table 4, off diagonals for percentages).

Cross-Informant Analyses

Results for cross-informant analyses examining the correspondence between by *parent-rated symptom groups* and *teacher-rated functioning* suggest that there was a significant association between symptom and functioning for two of the domains: self-esteem, $\chi^2(1, N = 43) = 4.07$, $p < .05$; and overall, $\chi^2(1, N = 43) = 6.23$, $p < .05$ (see Table 4, parent column). McNemar tests reveal significance in four of the six domains of functioning. In all four domains, there were significantly more youth who made reliable change in symptoms without reliable change in functioning than youth who made reliable change in functioning without reliable change in symptoms (see table for, parent column off diagonal).

The next set of cross-informant analyses examined the correspondence between *teacher-rated symptom groups* and *parent-rated functioning*. According to the Chi-square analyses, the associations between symptoms and functioning were not significant for any domain of functioning. McNemar tests indicated that youth were more likely to make reliable change in symptoms without change in peer relations (27%, $p < .03$) and self-esteem (30%, $p < .03$), than to make reliable change in these domains without change in symptoms (peer relations: 3%; self-esteem: 3%). For all other domains, the direction of discordance was not significant.

Discussion

Given the importance of impairment in the diagnosis of ADHD, as well as the breadth of impaired functional domains for children with the disorder, greater attention must be focused on the connection between change in symptoms and change in functioning over the course of treatment. The current study provides unique information about this connection in children with ADHD who received empirically-supported treatments in the school setting. Overall, both group and individual-level analyses indicate

that there are statistically significant correspondence between reliable change in symptoms and reliable change in functioning. However, the individual analyses reveal that there is a substantial minority (up to 40%) for whom there is change in one dimension without change in the other. Results and their implication are discussed.

Concordance

The results of the group-level analyses suggest that there is considerable correspondence between reliable change in symptoms and reliable change in functioning for parent ratings, teacher ratings, and cross-informant ratings (see Figs. 1, 2, and 4). Namely, symptom improvers evidenced a significant reduction in multiple domains of impairment across time, whereas symptom no-changers did not make significant change in functioning over time. This finding is consistent with a previous study that found that, as a group, children with ADHD who made reliable change in symptoms achieved reliable improvement in more functional domains than the group of symptom no-changers (Karpenko et al., 2009). Taken together, these data offer optimism that when high quality treatments are implemented, they have the potential to impact multiple domains of functioning as viewed by multiple informants.

It is important to keep in mind, however, that the MANOVA analyses examine changes in functioning at the group level. The second set of analyses directly examines the relation between reliable change in symptoms and reliable change in functioning (Jacobson & Truax, 1991) at the individual child level. Because it is often a child's functional impairments (e.g., decline in grades, fighting with peers), rather than symptoms, that result in a referral for services, it is important to determine the percentage of children who make reliable change in both outcome indicators, as well as those who make change in one indicator but not the other.

The individual-level analyses also indicated that there is a relationship between reliable change in symptoms and functioning on both parent and teacher reports. The greatest degree of correspondence was observed between symptoms and global ratings of functioning, where 50–74% of symptom improvers achieved reliable change in overall functioning (see Table 4, parenthetical percentages). The correspondence between symptoms and the domain-specific indicators of functioning was significantly lower, with one exception; teacher-rated symptoms and teacher-rated classroom functioning (62% correspondence). This relation is not surprising given that many ADHD symptoms (e.g., interrupts, out of seat, difficulty engaging in activities quietly) are directly related to classroom rules (raise hand to speak, remain in seat, work quietly).

Importantly, for all other specific domains of functioning, the percentage of symptom improvers who achieved

reliable improvement in functioning never exceeded 50% (see Table 4 parenthetical percentages). Thus, the majority of children who are classified as “treatment successes” according to symptom ratings would not be in that category according to functioning ratings. This is consistent with the findings that less than 20% of youth who demonstrated reliable change on the CBCL or YSR, also demonstrated reliable change on the CAFAS (Rosenblatt & Rosenblatt, 2002). Taken together, these findings suggest that the extant treatment efficacy data that are based primarily on change in symptoms may actually overstate the likelihood of treatment success for some children, particularly if functional impairments are most meaningful to families and teachers.

Discordance

Despite the above indicators of correspondence, McNemar tests across parent and teacher ratings indicated that there was a sizable minority of children who made reliable change in symptoms but not functioning (12–40% depending on the domain). In addition, there was a subset of children who achieved reliable change in functioning, but not symptoms (up to 16% depending on the domain). Further, the breakdown of these discordances by domain has important implications for treatment programming.

For example, in the peer domain, (consistent across parent and teacher report), approximately 25% of children experienced change in symptoms without experiencing a change in peer relations. This is consistent with the findings of the MTA study. Namely, treated children (across all four treatment conditions) made significant improvement in ADHD symptoms, yet all remained impaired in their peer relations according to sociometric assessment methods (Hoza et al., 2005), even those who had received intensive intervention focused on social relations.

In this study, very few children (3% by parent report to none by teacher report) experienced change in peer relationship without a change in symptoms. This percentage is substantially lower than that found in Karpenko et al. (2009), where 49% (by teacher report) to 52% (by parent report) of symptom no-changers made improvements in social skills. Arguably, peer sociometric assessment methods offer a different perspective on peer relations than do adult-rated social skills; however, the latter finding highlights that the relation between symptoms and domain-specific impairment may be a function of the intervention provided. For example, two-thirds of the MTA-treated children in the dataset analyzed by Karpenko et al. (2009) received intensive social skills training in the context of an 8-week summer treatment program, whereas children in the current study did not receive any social skills training or peer-focused intervention. Thus, future examination of the

correspondence between symptoms and domain-specific functioning should explore the extent to which the correspondence varies by the focus of the intervention.

Furthermore, others have documented that several symptoms associated with ADHD are viewed as annoying to other children (Pelham & Bender 1982) and likely to affect peer acceptance (Lopez-Williams et al., 2005). Thus, it could be argued that a reduction in these symptoms would be associated with an improvement in peer relations. That virtually none of the children in this study were able to improve peer relationships without a reduction in their ADHD symptoms supports the claim that the symptoms do interfere with peer relationships. The finding that very few children may reliable change in peer relations, in general, however, also suggests that the existing empirically-supported psychosocial interventions (i.e. behavioral parenting programs and behavioral classroom interventions) are likely insufficient to address the peer relationship problems associated with ADHD. Thus, new treatments and modifications are warranted to better address the mechanisms underlying social difficulties in ADHD.

A similar concern emerges with academic functioning and child–teacher relations, where nearly 40% of cases made reliable improvement in symptoms without improvements in academics or child–teacher relations. Given the connection between academic functioning and later life success, as well as the importance of student–teacher relationships in maintaining student engagement in schools (e.g., Klem & Connell, 2004), it is important that researchers and clinicians not become complacent when symptom change is achieved. It will be important to assess these other ecologically-relevant indicators to ensure that a change in symptoms has led to meaningful change in functional areas that have a long-term impact on adjustment. Such discordances in treatment outcome could inform treatment planning, including the decision to increase the dose of an intervention, to enhance domain- or setting- specific interventions, and to modify or terminate existing interventions.

Finally, there was a subsample (16%) of youth who made reliable change in parent–child relations without making reliable change in symptoms. On one hand, this finding indicates that a subset of parents may experience an improvement in this critical domain of functioning, and thus consider treatment clinically meaningful, without reporting reliable change in symptoms of ADHD. On the other hand, it is equally noteworthy that in both this study and Karpenko et al. (2009), less than half of the children who demonstrated reliable change in symptoms also demonstrated reliable change in parent–children relations and family functioning. As mentioned by Karpenko et al. (2009), this pattern may reflect the challenges associated with modifying coercive-family interactions that have become automated and ingrained within many families

with children with ADHD and disruptive behavior problems. However, this study also found that the parents of symptom improvers had significantly more contact hours with the clinician than symptom no-changers. Taken together, these findings suggest that on-going treatments (beyond a typical behavioral parenting program) or booster sessions are likely necessary to achieve optimal outcomes.

Cross-Informant Analyses

This is the first study to simultaneously examine concordance within informant and across informant. Not surprisingly, for some domains (i.e., peer relations, family functioning, classroom functioning) the within-informant correspondence is higher than the cross-informant (see Table 4). This likely reflects the uniqueness of the environments in which parents and teachers observe the children. However, interestingly, for other domains (e.g., parent–child relations, academic functioning), the cross-informant correspondence is comparable to the within-informant correspondence. This offers greater credibility to the results by reducing the likelihood that the associations found were inflated as a function of within source variance. Given that observed improvements in child functioning (e.g., academic performance) are likely to affect adult's perceptions of intervention effectiveness and treatment decision-making (e.g., to continue or discontinue the intervention, whether more intensive intervention is needed or not), it is important for research to begin to explicate the extent to which change in one informant's perceptions corresponds with change in another's.

Limitations

Several potential limitations should be discussed. First, the sample sizes for parent and teacher reports were not equal. Although *t*-test analyses indicate that the participants with parent data do not differ significantly on many important variables from those without parent data, this inequality may affect conclusions drawn. Second, some may argue that using a single item to capture a functional domain is a limitation. However, the IRS is frequently used to assess functional impairment in children with behavioral disorders (e.g., Evans, Langberg, Raggi, Allen, & Buvinger, 2005; Owens et al., 2008; Waxmonsky et al., 2008), and the IRS has demonstrated good psychometric properties (Fabiano et al., 2006). Indeed, because the IRS is short, feasible, and useful, it lends itself to use in real-world clinical settings, providing rich information about functioning across multiple relevant domains to aid in diagnostic decisions, treatment planning and services evaluation. Third, critics might argue that the current study should have used the two-criterion methodology for

determining CS change instead of using only RCI methodology. There are pros and cons to each of these methodologies, which have been outlined in the literature (Ogles, Lambert, & Sawyer, 1995; Tingey et al., 1996). Because ADHD is a chronic disorder, achieving normalization in both symptoms and functioning within a 1-year period may not be realistic. For these reasons, the authors believed that greater utility (i.e. results that may be applicable to a larger number of youth with the disorder) would be achieved by examining reliable change rather than CS change.

Implications

In the increasingly consumer-driven environment of mental health services, it is paramount that researchers examine factors most valuable to the consumers. In child mental health care, it is often functional impairment that causes the most distress and leads parents and teachers to refer a child to treatment. Further, recent work documents that symptoms are a related, but distinct construct from functional impairment (Gathje, Lewandowski, & Gordon, 2008). Indeed, the use of both symptoms and functioning in the diagnostic determination of the disorder, results in dramatically different identification rates than the use of symptoms only. By examining the reductions in both impairment and symptoms, researchers and practitioners can more fully understand the impact that interventions have on children and their families.

The current study has several important implications for both clinicians and researchers.

First, this study, along with a few other emerging studies in this area, confirms that reliable change in functioning *can* and *does* occur when there is no reliable change in symptoms (Kazdin, 2001, 2008). If symptom change alone was considered the indicator of treatment success, then a sizable percentage of children (up to 25%) could be deemed “treatment resistant” despite making reliable improvement in at least one critical domain of functioning. Similarly, a sizable number of children could be deemed “treatment successes” without having made gains in domains that are critical to healthy adjustment later in life (e.g., peer relations; Parker & Asher, 1987). As such, these data argue strongly for the use of impairment ratings over symptom ratings for evaluating treatment outcome in both research and practice. Indeed, in most cases, when impairment ratings improved, so did symptoms; yet the converse was less likely.

Second, this conclusion underscores the argument made by others in recent years about the importance of using multiple measures and multiple methods when examining meaningful outcomes (De Los Reyes & Kazdin, 2006; Kazdin, 1999, 2008). However, as the research community

creates systems for prioritizing treatment outcome measures, it is critical that these systems incorporate indicators that are most meaningful to consumers, as well as indicators that are practical under real-world service conditions. We would argue that the IRS represents a measurement tool that is feasible with regard to both time and money, and that produces information about functioning across a variety of domains that are relevant to both clinic-based and school-based services (Fabiano et al., 2006).

Third, the results highlight the importance of ongoing assessment of both symptoms and functioning throughout treatment. Without an infrastructure and process for assessment of both, inaccurate decisions may be made about treatment intensity, location, and termination. Similarly, those who are not responding in one area (e.g., symptom change) may also be experiencing a worsening in other domains even within a rather short period of time, as evidenced by Figs. 1, 2, and 3. The pattern of functioning associated with symptom no-changers stresses the need to examine and establish treatment algorithms that use data-driven decision-making throughout treatment and that attend to both dose and mode of treatment. Research examining treatment sequencing and dose is beginning to emerge (e.g., Pelham et al., 2008). It will be critical that the results of such work be communicated to both research and practice communities. Finally, these data also speak to the need for future research that examines the profiles of children who display no symptom change, or symptom deterioration (i.e., both pre-treatment demographic profiles, as well as treatment engagement and treatment response profiles), as well as potential moderators of positive treatment outcome.

Not only is functional impairment at the heart of the diagnostic classification for all mental health disorders, but it is also the driving force behind referrals to treatment. With this in mind, it seems presumptuous for researchers to conceptualize clinically meaningful and reliable change on the basis of symptomatic change alone. Results of the current study provide a valuable contribution to the scant research on the relation between reliable change in symptoms and important functional domains in children with ADHD. Although findings revealed some relation between changes in symptoms and functioning, it is notable that less than 50% of improvers in symptoms of ADHD had reliable improvement in the six domains of functioning. The present study also extended past research in the examination of clinically meaningful change using both within- and cross-informant analyses and by examining the correspondence between symptoms and functioning at the group and individual level. Taken together, the results support the need for including multiple informants, and multiple measures in the measurement of treatment outcome in order to capture clinically meaningful and reliable change in treatment.

References

- Achenbach, T. M. (1991a). *Manual for the child behavior checklist/4–18 and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the youth self report and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Angold, A., Costello, E. J., Farmer, E. M. Z., Burns, B. J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 129–137.
- Barkley, R. A. (1997). *Defiant children: A clinician's manual for assessment and parent training* (Vol. 2). New York: Guilford Press.
- Barkley, R. A. (2006). *Attention-deficit/hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). New York: Guilford Press.
- Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2006). Young adult outcome of hyperactive children: Adaptive functioning in major life activities. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 192–202.
- De Los Reyes, A., & Kazdin, A. E. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review*, 113, 554–583.
- Evans, S. W., Langberg, J., Raggi, V., Allen, J., & Buvinger, E. C. (2005). Development of a school-based treatment program for middle school youth with ADHD. *Journal of Attention Disorders*, 9, 343–353.
- Evans, S. W., & Youngstrom, E. (2006). Evidence-based assessment of attention-deficit/hyperactivity disorder: Measuring outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 1132–1137.
- Fabiano, G. A., Pelham, W. E., Coles, E. K., Gnagy, E. M., Chronis-Tusciano, A., & O'Connor, B. C. (2009). A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. *Clinical Psychology Review*, 29, 129–140.
- Fabiano, G. A., Pelham, W. E., Waschbusch, D. A., Gnagy, E. M., Lahey, B. B., Chronis, A. M., et al. (2006). A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology*, 35, 369–385.
- Gathje, R. A., Lewandowski, L. J., & Gordon, M. (2008). The role of impairment in the diagnosis of ADHD. *Journal of Attention Disorders*, 11, 529–537.
- Gordon, M., Antshel, K., Faraone, S., Barkley, R., Lewandowski, L., Hudziak, J. J., et al. (2006). Symptoms versus impairment: The case for respecting DSM-IV's criterion D. *Journal of Attention Disorder*, 9, 465–475.
- Hodges, K. (1994). *The child and adolescent functional assessment scale*. Ypsilanti, MI: Eastern Michigan University Psychology Department.
- Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven, CT: Yale University Department of Sociology.
- Hoza, B., Gerdes, A. C., Mrug, S., Hinshaw, S. P., Bukowski, W. M., Gold, J. A., et al. (2005). Peer-assessed outcomes in the Multimodal Treatment Study of children with attention deficit hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology*, 34, 74–86.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.

- Karpenko, V., Owens, J. S., Evangelista, N. M., & Dodds, C. (2009). Clinically significant symptom change in children with attention-deficit/hyperactivity disorder: Does it correspond with reliable improvement in functioning? *Journal of Clinical Psychology, 65*, 76–93.
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332–339.
- Kazdin, A. E. (2001). Almost clinically significant ($p < .10$): Current measures may only approach clinical significance. *Clinical Psychology: Science and Practice, 8*(45), 5–462.
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*, 146–159.
- Kelley, M. L. (1990). *School-home notes: Promoting children's classroom success*. New York: Guilford Press.
- Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*, 262–273.
- Lopez-Williams, A., Chacko, A., Wymbs, B. T., Fabiano, G. A., Seymour, K. E., Gnagy, E. M., et al. (2005). Athletic performance and social behavior as predictors of peer acceptance in children diagnosed with attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders, 13*, 173–180.
- MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*, 1073–1086.
- Ogles, B. M., Lambert, M. J., & Sawyer, J. D. (1995). Clinical significance of the National Institute of Mental Health treatment of depression collaborative research program data. *Journal of Consulting and Clinical Psychology, 63*, 321–326.
- Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*, 421–446.
- Owens, J. S., Murphy, C. E., Richerson, L., Girio, E. L., & Himawan, L. K. (2008). Science to practice in underserved communities: The effectiveness of school mental health programming. *Journal of Clinical Child and Adolescent Psychology, 37*, 434–447.
- Owens, J. S., Richerson, L., Beilstein, E. A., Crane, A., Murphy, C. E., & Vancouver, J. B. (2005). School-based mental health programming for children with inattentive and disruptive behavior problems: First year treatment outcome. *Journal of Attention Disorders, 9*, 261–274.
- Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102*, 357–389.
- Pelham, W. E. (2002). Attention deficit hyperactivity disorder: Diagnosis, assessment nature, etiology, and treatment (unpublished manual).
- Pelham, W. E., & Bender, M. E. (1982). Peer relationships in hyperactive children: Description and treatment. In K. D. Gadow & I. Bialer (Eds.), *Advances in learning and behavioral disabilities* (Vol. 1, pp. 365–436). Greenwich, CT: JAI Press.
- Pelham, W. E., & Fabiano, G. A. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology, 37*, 184–214.
- Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 449–476.
- Pelham, W. E., Fabiano, G., Waxmonsky, J., Verley, J., Gnagy, E. M., Greiner, A., Bhatia, I., & Tresco, K. (June 2008). *Adaptive treatments for children with ADHD*. Poster presented at the institute of educational sciences, third annual research conference, Washington, DC.
- Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 210–218.
- Pelham, W. E., Wheeler, T., & Chronis, A. (1998). Empirically supported psychosocial treatments for ADHD. *Journal of Clinical Child Psychology, 27*, 190–205.
- Rosenblatt, A., & Rosenblatt, J. A. (2002). Assessing the effectiveness of care for youth with severe emotional disturbances: Is there agreement between popular outcome measures? *The Journal of Behavioral Health Services & Research, 29*, 259–273.
- Sheridan, S. M., Kratochwill, T. R., & Bergan, J. R. (1996). *Conjoint behavioral consultation: A procedural manual*. New York: Plenum Press.
- Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., et al. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 168–179.
- Tingey, R. C., Lambert, M. J., Burlingame, G. M., & Hansen, N. B. (1996). Clinically significant change: Practical indicators for evaluation psychotherapy outcome. *Psychotherapy Research, 6*, 144–153.
- Waxmonsky, J., Pelham, W. E., Gnagy, E., Cummings, M. R., O'Connor, B., Majumdar, A., et al. (2008). The efficacy and tolerability of methylphenidate and behavior modification in children with attention-deficit/hyperactivity disorder and severe mood dysregulation. *Journal of Child and Adolescent Psychopharmacology, 18*, 573–588.
- Weller, E. B., Weller, R. A., Teare, M., & Fristad, M. A. (1999). *Parent version-children's interview for psychiatric syndromes (P-ChIPS)*. Washington, DC: American Psychiatric Press.
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82*, 50–59.