

# Parent and Teacher Ratings on the IOWA Conners Rating Scale

Daniel A. Waschbusch · Michael T. Willoughby

Published online: 11 September 2007  
© Springer Science + Business Media, LLC 2007

**Abstract** The IOWA Conners Rating Scale is a widely used brief measure of inattentive-impulsive-overactive (IO) and oppositional-defiant (OD) behavior in children. This study examined the psychometric properties of this measure when completed by mothers and teachers. Results of confirmatory factor analyses indicated that a three-factor solution, conforming to current DSM-IV formulations of the disruptive behavior disorders, provided a better fit to the observed data than the currently used two-factor model, in which no distinction is made between inattentive and hyperactive-impulsive behaviors. Both new and currently used scale scores had good internal consistency and test-retest reliability and showed that boys' scores were significantly higher than girls' scores. Results held for both mother and teacher ratings. Clinical cutoff scores were proposed and performed reasonably well to screen for ADHD and ODD. Results support the IOWA Conners as a screening measure for the disruptive behavior disorders or as a tool for monitoring treatment response.

**Keywords** ADHD · ODD · Screening · Treatment evaluation · Behavior ratings

---

D. A. Waschbusch (✉)  
Center for Children and Families,  
Departments of Pediatrics and Psychology,  
University at Buffalo–State University of New York,  
Diefendorf Hall, Room 106, 3435 Main Street, Bldg 20,  
Buffalo, NY 14214, USA  
e-mail: dw35@buffalo.edu

M. T. Willoughby  
Frank Porter Graham Child Development Institute,  
University of North Carolina,  
Sheryl Mar Bldg., CB #8185,  
Chapel Hill, NC 27599-8185, USA  
e-mail: willoughby@unc.edu

Externalizing behaviors are among the most common mental health problems among elementary school children (Lahey et al. 1999). Externalizing behavior problems are typically divided into two classes: (1) inattentive, impulsive, overactive (IO) behaviors that are the primary characteristics of Attention Deficit/Hyperactivity Disorder, and (2) oppositional, defiant, rule breaking (OD) behaviors that are the primary characteristics of oppositional defiant disorder and conduct disorder (American Psychiatric Association 2000). Numerous studies demonstrate that IO and OD behaviors are distinct but highly associated constructs, whether they are measured continuously or categorically (Hinshaw 1987; Waschbusch 2002), and that both sets of behaviors are significantly associated with academic difficulties, impairment in peer relationships, and conflicts with parents and teachers (Quay and Hogan 1999).

Empirically supported assessment is vital to understanding and treating externalizing problems in children and standardized behavior rating scales are an essential component of empirically supported assessment (McMahon and Frick 2005; Pelham et al. 2005a, b). One commonly used measure is the Inattention/Overactivity with Aggression (IOWA) Rating Scale. The IOWA was first developed by Loney and Milich (1982) by selecting items (see Table 1) from the abbreviated Conners Teacher Rating Scale (Conners 1969) that best differentiated IO behaviors from OD behaviors in children. The result of this work was a ten item measure that included a five item inattention-impulsivity-overactivity (IO) scale and a five item oppositional-defiance (OD) scale. The reliability and validity of the IOWA scales were supported in the initial study (Loney and Milich 1982), and subsequent work has continued to find high reliability and validity of the IOWA when completed by teachers (Atkins et al. 1988, 1989; Johnston and Pelham 1986; Milich and Fitzgerald 1985; Milich and Landau 1988; Nolan and

**Table 1** Polychoric correlations for mother and teacher-reported IOWA Conners items

Items and Scale	1	2	3	4	5	6	7	8	9	10
Inattentive-impulsive-overactive (IO) items										
1. Fidgeting	–	0.79	0.77	0.87	0.73	0.60	0.54	0.61	0.65	0.63
2. Hums and makes other odd noises	0.62	–	0.77	0.74	0.64	0.66	0.59	0.70	0.71	0.69
3. Excitable, impulsive	0.68	0.63	–	0.71	0.55	0.68	0.65	0.68	0.66	0.64
4. Inattentive, easily distracted	0.68	0.50	0.70	–	0.88	0.64	0.56	0.61	0.68	0.69
5. Fails to finish things he or she starts (short attention span)	0.62	0.47	0.57	0.85	–	0.57	0.47	0.59	0.62	0.66
Oppositional-defiant (OD) items										
6. Quarrelsome	0.47	0.44	0.59	0.60	0.55	–	0.77	0.86	0.89	0.85
7. Acts ‘smart’	0.47	0.54	0.57	0.57	0.52	0.76	–	0.73	0.74	0.74
8. Temper outbursts - behavior explosive and unpredictable	0.43	0.49	0.58	0.56	0.48	0.75	0.63	–	0.89	0.84
9. Defiant	0.50	0.50	0.60	0.58	0.53	0.82	0.74	0.82	–	0.93
10. Uncooperative	0.49	0.50	0.60	0.62	0.58	0.78	0.70	0.78	0.87	–

Items 1 through 5 are the inattentive-impulsive-overactive (IO) items and items 6 through 10 are the oppositional defiant (OD) items. Mother and teacher values appear below and above the diagonal, respectively; Mother  $N=728$ ; Teacher  $N=1517$ ; All  $p<0.0001$

Gadow 1994). In addition, normative data have been published for teacher ratings on the IOWA (Pelham et al. 1989). In the years since this work, the IOWA has become one of the most commonly used measures of IO and OD in children. Indeed, the publication outlining normative data for teacher ratings on the IOWA (Pelham et al. 1989) has been cited in more than 100 studies, including recent treatment studies (e.g., Remschmidt et al. 2005), laboratory studies (e.g., Oosterlan et al. 2005), family studies (e.g., Brent et al. 2004), and assessment studies (e.g., Gadow et al. 2004; Halperin et al. 2003).

The continued popularity of the IOWA may seem surprising given that many other empirically supported rating scales have been published in the nearly 25 years since the IOWA was first developed (see McMahon and Frick 2005; Pelham et al. 2005a, b for a review). The continued use of the IOWA is particularly surprising considering that many of the more recent rating scales for assessing externalizing behavior in children have the advantage of being indexed to the current version of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association 2000) whereas the IOWA does not. Despite this important disadvantage, the IOWA continues to be widely used because it provides a number of relatively unique advantages not offered by many other rating scales.

First, the IOWA can be freely used by anyone with appropriate qualifications, whereas many other rating scales for evaluating IO and OD behaviors in children require a fee for each administration. Second, the IOWA is exceptionally brief. This makes it especially useful in situations when assessments are being administered repeatedly within short periods of time (e.g., in treatment studies), when conducting screening assessments on a large number of

children, or when a large number of constructs are being measured at once. Third, the IOWA has been widely and consistently used in the same form over a number of years and across a wide variety of studies, providing the ability to compare results of current research and clinical efforts with past efforts in a relatively direct and meaningful way.

At the same time, there are at least three disadvantages associated with the IOWA rating scale. First, in contrast to the considerable body of evidence supporting the psychometric properties of the IOWA when completed by teachers, little or no research has examined the psychometric properties of the scale when completed by mothers. This is a serious limitation because many children are referred for treatment by their mothers and because mother ratings are often used to evaluate treatment response (e.g., Elgar et al. 2003; Pelham et al. 2005a, b; Remschmidt et al. 2005). Second, the cross-informant equivalence of mother and teacher use of the IOWA is unknown. In general, the correlation between parent and teacher ratings of externalizing is low (average  $r=0.21$ , as per Achenbach et al. 1987), despite the fact that both provide valid information about externalizing behavior in children (e.g., Hart et al. 1994; Loeber et al. 1989). An improved understanding of cross-informant agreement would better inform questions about whether discrepancies between parent and teacher reports of child behavior using the IOWA reflect true behavior differences that occur across home and school contexts or instead reflect differential functioning of the scale across informants. Third, there is growing consensus that hyperactive/impulsive (HY) and inattentive (IN) behaviors are conceptually and empirically distinct (e.g., Barkley 1997; Lahey et al. 1988; Milich et al. 2001; Pillow et al. 1998), but the IO scale of the IOWA Conners conflates these two dimensions of behavior.

The purpose of this study was to examine the factor structure of the IOWA when employed by mothers and teachers and to examine the psychometric properties and performance of the factors. More specifically, five hypotheses were examined. First, it was hypothesized that a three factor model of the IOWA (consisting of hyperactive/impulsive, inattentive, and oppositional defiant factors) would provide a better fit to the data than the more commonly used two factor model (consisting of an inattentive-overactive-impulsive (IO) and oppositional defiant (OD) factors). Second, it was hypothesized that the factor structure of the IOWA would be equivalent across mothers and teachers. Third, it was hypothesized that both the three factor and two factor conceptualizations of the IOWA would demonstrate adequate reliability (internal consistency and test-retest) for both mothers and teachers. Fourth, it was hypothesized that factors would differ as a function of both grade and gender, as has been found in previous research examining teacher ratings on the IOWA (Pelham et al. 1989). Fifth, it was hypothesized that screening cutoffs based on the 90th percentile of the distribution would be useful for identifying children meeting diagnostic criteria for ADHD and ODD, as per the full set of DSM-IV symptoms and corresponding impairment criteria.

## Method

### Participants

Participants were mothers and/or teachers of 1,554 children who were students in one of seven public schools in a single school district in northern Nova Scotia, Canada. The participating school district included 58 elementary schools which served approximately 13,000 elementary students (Nova Scotia Department of Education 2003). The students ranged in age from 5 to 12 (mean=8.13; SD=1.94) and consisted of 809 boys (52.1%) and 745 girls (47.9%). Ethnic and racial information of participants was not collected (at the request of the school board), but the schools served communities that were over 95% Caucasian (Nova Scotia Department of Finance 2003). Of the 1,554 students, complete teacher data on the IOWA were available for 1,517 students, complete mother data on the IOWA were available for 728 students, and complete teacher and mother data on the IOWA were available for 711 students.

### Measures

*IOWA Conners Rating Scale* The IOWA Conners Rating Scale (Loney and Milich 1982; Pelham et al. 1989) consists of ten items (see Table 1) each of which is evaluated using

a four point Likert scale with the following anchors: not at all (0); just a little (1); pretty much (2); and very much (3). The first five items on the IOWA are designed to measure inattentive-impulsive-overactive (IO) behaviors and the second five items are designed to measure oppositional-defiant (OD) behaviors. Included in the IO scale are three items (items 1, 2 and 3) that measure hyperactive/impulsive behaviors and two items (items 4 and 5) that measure inattentive behaviors. The IOWA was completed by mothers and teachers.

*Assessment of Disruptive Symptoms DSM-IV (ADS-IV)* The ADS-IV is a rating scale designed to measure ADHD and ODD (as defined in the DSM-IV) in elementary school children (Waschbusch et al. 2003). The majority of items on the ADS-IV are symptoms of ADHD and ODD taken directly from the DSM-IV, with minor wording changes to make them more concise and appropriate for a rating scale format. Specifically, the word “often” was removed from all symptoms, as others have done (Burns et al. 1997), and some items were simplified (e.g., “Often blurts out answers before questions have been completed” was changed to “Blurts out answers before questions have been finished”). Each symptom is rated on a 0 (“much less than other children”) to 4 (“much more than other children”) Likert scale to indicate the extent to which the child expresses the symptom relative to others of the same age and gender. Symptoms were counted as present if they were rated 3 (“more than other children”) or 4 (“much more than other children”).

Following the nine ADHD-inattention symptoms, impairment resulting from symptoms is assessed by requesting teachers and parents rate the degree to which inattention symptoms cause the child problems, with possible responses ranging from (0) “no problems” to (4) “very severe problems.” Teachers rate the extent to which inattention caused problems at school, whereas parents rate the extent to which inattention caused problems at home, school, or other places (e.g., playground). Parents also estimate the age of onset of inattention problems. These questions were also asked following the nine ADHD-hyperactive/impulsive symptoms and following the eight ODD symptoms. The ADS-IV was requested from mothers and teachers. Complete teacher ratings on the ADS-IV were available for 1482 children (97.7% of the children with complete teacher rating on the IOWA), and complete mother ratings on the ADS-IV were available for 708 children (97.3% of the children with complete mother ratings on the IOWA).

The ADS-IV was used to form ADHD and ODD groups by applying DSM-IV symptom count and impairment criteria. Children were assigned to the ADHD group if: (a) they were rated as having at least six ADHD-inattention

symptoms and were rated as having severe or very severe impairment from the inattention symptoms; (b) they were rated as having at least six ADHD-hyperactive/impulsive symptoms and were rated as having severe or very severe impairment from hyperactive/impulsive symptoms; or (c) they met both a and b criteria. Similarly, children were assigned to the ODD group if they were rated as having at least four ODD symptoms and were rated as having severe or very severe impairment from ODD symptoms. These categorical scores have been shown to be significantly associated with similar scores derived from other rating scales (the Disruptive Behavior Disorders Rating Scale; Pelham et al. 1992) and from structured diagnostic interviews (i.e., the Diagnostic Interview Scale for Children; NIMH-DISC Editorial Board 1999), as reported elsewhere (Waschbusch et al. 2007, 2003, 2006).

### Procedure

Data were collected as part of the pre-intervention evaluation of the Behavior Education Support and Treatment (BEST) school intervention program. The BEST project was designed to prevent and treat disruptive behavior in elementary school settings using behavioral strategies delivered at universal, targeted, and clinical levels (see Waschbusch et al. 2005 for details). All procedures and measures were approved by a university Human Ethics Review Board and by the participating schools and school district and informed consent was obtained from parents and teachers. Schools were recruited by contacting principals and providing them information about the project through presentations and written materials. Principals then met with their staff and subsequently contacted the project coordinator if their school wished to participate. Seven schools volunteered to participate. Three of these seven schools were randomly assigned to the intervention condition and the remaining four participated as controls.

Approximately 6 weeks after the start of the school year, but before the start of the intervention, teachers and parents of students in participating schools were asked to complete a packet of rating scales, including those used in this study. All raters were told that completing the ratings was voluntary and that their responses would be confidential. Homeroom teachers were given the option of taking an in-service day to complete ratings on students in their homeroom classrooms, and all teachers ( $n=66$ ) elected to do so. Therefore, ratings by homeroom teachers were completed for nearly all students in participating schools, although teachers of 37 students (representing less than 1% of the sample) returned packets without complete data on the primary measure of interest in this study (the IOWA) and were therefore excluded from the analyses involving

teacher ratings. Mothers of 728 children (46.8% of the eligible sample) returned packets with complete IOWA data. Of the 728 children with complete mother ratings, 711 (97.7%) also had complete teacher ratings. The remaining 826 mothers (53.2%) did not return complete IOWA ratings and were excluded from analyses of mother ratings. This rate of parental participation is consistent with other school-based assessments (e.g., Airaksinen et al. 2004; Frissell et al. 2004; Gottfredson and Gottfredson 2001; McGrew and Gilman 1991; Wolraich et al. 2004). Comparison of students with and without complete mother ratings showed that groups did not differ on age, sex, or teacher ratings of behavior.

These procedures were repeated at the end of the school year, approximately seven months later. The ratings collected at the end of the school year (time 2) were used to estimate test-retest reliability. Only students in control schools were used in test-retest analyses to remove any confounding effects of the intervention. There were 706 students in control schools with teacher IOWA data at time 1. Of these, 640 (90.7%) also had teacher IOWA data at time 2. There were 357 students in control schools with mother IOWA data at time 1. Of these, 182 (51.0%) also had mother IOWA ratings at time 2.

### Data Analysis

This study was organized around five research questions that correspond with the hypotheses described earlier. The first question investigated the dimensionality of the IOWA Conners items. Specifically, the relative fit of two- and three-factor models were tested, both of which were specified a priori. The second question investigated whether the psychometric properties of the IOWA Conners were equivalent across informants. The third question investigated the test retest reliability of the IOWA Conners scale scores. The fourth question provided *preliminary* normative data for scale scores by child gender and grade, separately for mother and teacher informants, including 90th percentile cutoff scores for purposes of screening. The fifth question evaluated the proposed screening cutoffs by comparing them to parent and teacher-ratings of DSM-IV symptoms.

The first two questions were addressed using confirmatory factor analyses (CFAs), while the remaining three questions were accomplished using univariate statistics including bivariate correlations, as well as computation of recommended diagnostic statistics (Kessel and Zimmerman 1993). All CFA models were fit using Mplus version 4.00 using robust weighted least squares (WLSMV) estimation to accommodate the fact that IOWA items were rated using a four-point Likert scale, with the majority of respondents relying on the two lowest categories (Muthén et al. 1997).

A recent simulation study indicated that the WLSMV estimation outperformed standard WLS estimation for the types of data, models, and sample sizes that are considered here (Flora and Curran 2004). Given the known dependence between sample size and chi-square model fit statistics, the comparative fit index (CFI), root mean squared error of approximation (RMSEA), and the weighted root mean square residual (WRMR) were used as indices of *absolute* model fit. Values of  $CFI \geq 0.95$ ,  $RMSEA \leq 0.05$ , and  $WRMR \leq 1.0$  are indicative of good model fit, whereas values of CFI between 0.90 and 0.95, RMSEA between 0.06 and 0.07, and WRMR 1.0–1.5 indicate only moderately good fit, for the models and sample sizes that are considered here (Hutchinson and Olmos 1998; Yu 2003). More central to this research was the *relative* fit of two versus three factor models. Questions of relative fit were addressed using scaled chi-square difference tests (see Appendix 5 in Muthén and Muthén 2004; Satorra and Bentler 1999). The reader should appreciate that the degrees of freedom for WLSMV models are estimated (see Muthén et al. 1997 for elaboration).

## Results

### Factor Structure

*Bivariate Correlations* Polychoric correlations for the 10 IOWA Conners items are summarized in Table 1, with mother-reported items below the diagonal and teacher-reported items above the diagonal.

*Mother Ratings* A synopsis<sup>1</sup> of fit for all CFA models is provided in Table 2. In terms of absolute fit, the two- and three-factor models fit the data moderately well. In terms of relative fit, the three-factor model provided a better fit to the data than did the two-factor model,  $\chi^2(2)=57.6$ ,  $p < 0.0001$ .

*Teacher Ratings* A parallel set of CFA models were estimated for teachers. In terms of absolute fit, the two-factor model fit the data rather poorly while the three-factor model fit the data moderately well. In terms of relative fit, the three-factor model provided a better fit to the data than did the two-factor model,  $\chi^2(2)=113.5$ ,  $p < 0.0001$ .

<sup>1</sup> Given the sampling design, students were nested within classrooms. To assess whether this hierarchical data structure affected any of the conclusions reported herein, sensitivity analyses were performed. Specifically, every CFA model was re-estimated as multilevel CFA model such that children were considered clustered in classrooms. Rather than specifying separate classroom and child level models, classroom was treated as a nuisance variable and used to obtain corrected test statistics and standard errors. None of the substantive conclusions changed once the hierarchical data structure was taken into account. These analyses are available upon request.

*Simultaneous Mother and Teacher Ratings* A final set of two- and three- factor CFA models were simultaneously estimated for the 711 children who had complete mother and teacher data on the IOWA (i.e., a total of four factors, two for each informant, versus a total of six factors, three for each informant, were simultaneously estimated). In terms of absolute fit, the two-factor model fit the data moderately well while the three-factor model now provided a good fit to the data, as indexed by all three fit indices. In terms of relative fit, the WLSMV-scaled chi-square difference tests once again confirmed that the three-factor model fit better than did the two-factor model,  $\chi^2(6)=82.1$ ,  $p < 0.0001$ .

### Informant Equivalence

Having established that a three-factor model provided superior fit to mother and teacher-reported data, both alone and in combination, the next test examined whether individual IOWA items were related to their respective latent factors in an identical manner across informants. This was accomplished by re-estimating the three-factor model simultaneously for both informants with equality constraints imposed on all factor loadings. Whereas establishing that a similar number of factors underlie both mother and teacher-reported data is known as configural invariance, the test of equivalence of factor loadings across informants is known as a test of weak invariance (Meredith 1993). Although this model appeared to fit the data reasonably well (see Table 2), a scaled chi-square difference test indicated that the constrained model fit the data significantly worse than did the model in which factor loadings were free to vary across informants,  $\chi^2(6)=28.1$ ,  $p=0.0001$ . Inspection of individual factor loadings across informants did not indicate any items that were noticeably different across informants. This suggested that the significance of the scaled chi-square difference tests was a function of the large sample size (and hence high power). Thus, a final three-factor model was tested, simultaneously for mother and teacher informants, and only required that a single item for each factor (item 2: Hums and makes other odd noises; item 5: Fails to finish things; item 8: temper outbursts) take on equal factor loadings. This is known as partial measurement invariance and is the minimally sufficient condition for establishing that a scale works equivalently across groups (Byrne et al. 1989). The fit of this model was once again consistent with the previous two models (see Table 2). More importantly, this restricted model fit equally well as the model without constraints on factor loadings (i.e., partial measurement invariance was supported),  $\chi^2(3)=1.7$ ,  $p=0.63$ .

As a result of establishing partial measurement invariance across raters, cross-informant correspondence for the

**Table 2** Synopsis of model fit for confirmatory factor analyses

Informant	Description	N	Model fit			
			$\chi^2(df)$	CFI	RMSEA	WRMR
Mother-only	2 factors	728	173.8 (19)	0.97	0.11	1.4
	3 factors	728	103.6 (21)	0.98	0.07	0.9
Teacher-only	2 factors	1517	325.2 (17)	0.98	0.11	2.0
	3 factors	1517	225.4 (19)	0.99	0.09	1.4
Mother + teacher	2 factors	711	217.4 (51)	0.98	0.07	1.2
	3 factors	711	133.4 (52)	0.99	0.05	0.9
	3 factor (all equal loadings)	711	146.5 (51)	0.99	0.05	1.0
	3 factor (three equal loadings)	711	126.5 (52)	0.99	0.05	0.9

All chi square test statistic  $p < 0.0001$

*df* Degrees of freedom, *CFI* comparative fit index, *RMSEA* root mean squared error of approximation, *WRMR* weighted root mean square residual

hyperactive-impulsive (HY), inattentive (IN), and oppositional defiant (OD) scores was examined using latent correlations, which were not attenuated by measurement error. As is summarized in Table 3, the latent and observed correlations between factors *within* informant and *across* informants were significantly different from zero. The relative improvement in cross-informant agreement using latent versus observed correlations is also evident by comparing corresponding values above and below the diagonal.

**Computing Scale Scores**

The preceding results demonstrated that the IOWA Conners items are best conceptualized as resulting from three factors (IN, HY, OD). This differs from current practice where IN and HY scores are aggregated to form a single inattentive-overactive scale (IO). Given the widespread use of the IOWA IO scale, it was included in all of the remaining analyses, along with the IN, HY, and OD scales that were supported by CFA models. Moreover, because a majority of the people who use the IOWA Conners will do so using the observed scores, the remaining analyses are based exclu-

sively on scale scores that represent the sum of the observed items (not latent variables). More specifically, based on the a priori model and accepted use of the IOWA, an IO score was computed by summing together items 1 through 5 and an OD score was computed by summing together items 6 through 10 (see Table 1 for items and item numbers). In addition, based on CFA results, the IO score was further divided into a hyperactive/impulsive (HY) score by summing items 1, 2 and 3, and into an inattention (IN) score by summing items 4 and 5. Means and standard deviations for these scales are summarized in Table 4.

**Reliability**

*Mother Ratings* Coefficient alpha was used to evaluate the internal consistency and Pearson correlations were used to evaluate test–retest reliability of the IOWA scales as completed by mothers. Results are summarized in Table 4.

*Teacher Ratings* The same reliability procedures were used to evaluate reliability of IOWA scales as completed by teachers. Results are summarized in Table 4.

**Table 3** Cross informant correlations

	1	2	3	4	5	6
Mother IN	–	0.65	0.60	0.51	0.47	0.35
Mother HY	0.82	–	0.62	0.34	0.42	0.38
Mother OD	0.69	0.74	–	0.28	0.31	0.31
Teacher IN	0.62	0.42	0.36	–	0.78	0.51
Teacher HY	0.58	0.53	0.40	0.93	–	0.60
Teacher OD	0.52	0.54	0.45	0.69	0.79	–

Latent and observed correlations appear below and above the diagonal, respectively;  $N=711$ ; all  $p < 0.001$

*IN* Inattentive, *HY* hyperactive-impulsive, *OD* oppositional-defiant

**Table 4** Means, standard deviations (SD), and reliability coefficients for IOWA scales as a function of informant

Scale	Mean	SD	Reliability coefficients	
			$\alpha$	$r_{1,2}$
Mother ratings				
Inattentive-overactive-Impulsive (IO)	3.64	3.23	0.86	0.80
Oppositional-defiant (OD)	3.10	3.28	0.91	0.69
Hyperactive-impulsive (HY)	2.23	2.01	0.78	0.79
Inattentive (IN)	1.40	1.54	0.85	0.76
Teacher ratings				
Inattentive-overactive-impulsive (IO)	3.34	3.58	0.90	0.79
Oppositional-defiant (OD)	1.03	2.37	0.90	0.75
Hyperactive-impulsive (HY)	1.82	2.14	0.85	0.78
Inattentive (IN)	1.52	1.71	0.87	0.73

IO scale was computed by summing items 1 through 5 (all items and item numbers are listed in Table 1). OD was computed by summing items 6 through 10. HY was computed by summing items 1 through 3. IN was computed by summing items 4 and 5

$\alpha$  Chronbach's alpha internal consistency reliability,  $r_{1,2}$  test-retest reliability

### Gender and Grade Differences

**Mother Ratings** Data from 728 mothers were used to examine grade and gender differences using a series of 2 (gender)  $\times$  3 (grade: K/1 vs 2/3 vs 4/5/6) ANOVAs<sup>2</sup>. As summarized in Table 5, there was a significant main effect of gender for each scale, but no main effects or interactions involving grade. Examination of means and standard deviations for the gender main effect (see Table 5) showed that boys had higher scores than girls on every scale, with differences characterized by small to medium effect sizes (Cohen's D: IO=0.29; OD=0.20; HY=0.24; IN=0.31).

**Teacher Ratings** Grade  $\times$  Gender ANOVAs were also used to examine data from the 1,517 teacher rating. As summarized in Table 5, there were significant main effects of gender for each scale, and a significant main effect of grade for HY and IN. Examination of means and standard deviations for the gender main effect (see Table 5) showed that boys had higher scores than girls on each scale, with differences characterized by small to medium effect sizes (Cohen's D: IO=0.38; OD=0.26; HY=0.36; IN=0.35). Tukey honestly significant difference (HSD) post hoc tests of the main effect of grade for HY showed that children in grades K/1 had significantly higher HY scores than children in other grades, but other groups did not differ (Grades K/1: mean=2.10, SD=2.25; Grades 2/3: mean=1.67, SD=2.20; Grades 4/5/6: mean=1.75, SD=2.00). These differences were characterized by small effect sizes (Cohen's D: K1/vs

2/3=0.20; K/1 vs 4/5/6=0.17; 2/3 vs 4/5/6=-0.05). Tukey HSD tests of the main effect of grade for IN showed that children in grades 4/5/6 had significantly higher IN scores than children in grades 2/3, but other groups did not differ (Grades K/1: mean=1.43, SD=1.63; Grades 2/3: mean=1.40, SD=1.73; Grades 4/5/6: mean=1.67, SD=1.74). These differences were characterized by small effect sizes (Cohen's D: K1/vs 2/3=0.02; K/1 vs 4/5/6=-0.14; 2/3 vs 4/5/6=-0.16).

### Screening Cutoffs

The 90th percentile was selected as the criteria for identifying children with elevated scores. In other words, children in the upper 10% of the distribution were identified as having significantly elevated scores. The 90th percentile was selected to provide a reasonable estimate of the prevalence of externalizing behavior problems while minimizing the chance of making false negative errors (i.e., failing to identify children who should be identified), which are arguably the most serious type of error when conducting screening assessments. Cut scores were computed separately for each scale and separately for boys and girls. The proposed scale cutoffs, defined as the minimal score to identify children at or above the 90th percentile, are summarized in Table 5 for mother and teacher ratings.

### Evaluation of Screening Cutoffs

**Overview** The proposed cutoffs for the IOWA scales were examined using a series of 2 $\times$ 2 chi-square analyses to compare the IOWA groups (IO, OD, HY, IN) to comparable

<sup>2</sup> Grade rather than age was used in these analyses to maintain consistency with past research on the IOWA (Pelham et al. 1989).

**Table 5** ANOVA results, means, standard deviations (SD), and cutoffs for identifying clinically elevated scores on the IOWA Conners

Scale	Boys			Girls			ANOVA results		
	Mean	SD	Cutoff	Mean	SD	Cutoff	Gender <i>f</i>	Grade <i>f</i>	Grade × gender <i>f</i>
<b>Mother ratings</b>									
IO	4.10	3.34	10	3.17	3.05	8	14.51***	0.63	1.68
OD	3.42	3.47	9	2.78	3.17	7	6.29*	1.44	0.95
HY	2.47	2.06	6	2.00	1.94	5	9.11**	1.11	0.01
IN	1.64	1.61	5	1.17	1.43	4	16.37***	1.66	0.55
<b>Teacher ratings</b>									
IO	4.00	3.84	10	2.63	3.13	7	52.72***	1.72	0.65
OD	1.33	2.65	5	0.71	1.97	3	23.23***	1.18	0.53
HY	2.19	2.32	6	1.41	1.84	4	49.22***	4.84**	0.14
IN	1.80	1.82	5	1.22	1.54	4	41.27***	3.92*	1.48

Proposed cutoffs are minimal scores for identifying children with significant elevations (defined as at or above the 90th percentile) on the scale. Degrees of freedom for mother ratings: Gender=1, 722, Grade=2, 722, Gender × grade=2, 722. Degrees of freedom for teacher ratings: Gender=1, 1,511, Grade=2, 1,511, Gender × grade=2, 1,511.

IO Inattentive/impulsive/overactive, OD oppositional-defiant, HY hyperactive/impulsive, IN inattentive. *ff* value in ANOVA

\**p*<0.05

\*\**p*<0.01

\*\*\**p*<0.001

ADS-IV groups. A variety of recommended diagnostic evaluation statistics were also computed from these analyses, including: sensitivity, specificity, positive predictive power, negative predictive power, overall percent correct, and kappa (Kessel and Zimmerman 1993). Only children with complete data on both the IOWA and on the ADS-IV were used for these analyses (mother ratings *n*=708, teacher ratings *n*=1,482).

*Mother Ratings* The 2×2 chi-square analyses showed: (a) IOWA IO (no vs yes) was significantly related to ADS-IV ADHD-any subtype (no vs yes),  $\chi^2(1)=247.59, p<0.001$ ; (b) IOWA OD (no vs yes) was significantly related to ADS-IV ODD (no vs yes),  $\chi^2(1)=223.26, p<0.001$ ; (c) IOWA HY (no vs yes) was significantly related to ADS-IV ADHD hyperactive/impulsive/combined types (no vs yes),  $\chi^2(1)=136.12, p<0.001$ ; (d) IOWA IN (no vs yes) was signifi-

**Table 6** Number (percent) of children in IOWA groups and ADS-IV groups as rated by mothers

	ADS-IV ADHD-any subtype		Total	Diagnostic statistics
	No	Yes		
<b>IOWA IO</b>				
No	638 (90.1%)	8 (1.1%)	646 (91.2%)	Sens=0.79; Spec=0.95 PPP=0.48; NPP=0.99 Overall=0.94; $\kappa=0.57$
Yes	32 (4.5%)	30 (4.2%)	62 (8.8%)	
Total	670 (94.6%)	38 (5.4%)	708 (100%)	
<b>IOWA OD</b>				
ADS-IV ODD				
No	627 (88.6%)	11 (1.6%)	638 (90.1%)	Sens=0.75; Spec=0.94 PPP=0.47; NPP=0.98 Overall=0.9; $\kappa=0.54$
Yes	37 (5.2%)	33 (4.7%)	70 (9.9%)	
Total	664 (93.8%)	44 (6.2%)	708 (100%)	
<b>IOWA HY</b>				
ADS-IV ADHD-HY and combined				
No	640 (90.4%)	4 (0.6%)	590 (83.3%)	Sens=0.81; Spec=0.93 PPP=0.27; NPP=0.99 Overall=0.93; $\kappa=0.37$
Yes	47 (6.6%)	17 (2.4%)	118 (16.7%)	
Total	687 (97.0%)	21 (3.0%)	708 (100%)	
<b>IOWA IN</b>				
ADS-IV ADHD-IN and combined				
No	648 (91.5%)	9 (1.3%)	657 (92.8%)	Sens=0.68; Spec = .95 PPP=0.37; NPP=0.99 Overall=0.94; $\kappa=0.45$
Yes	32 (4.5%)	19 (2.7%)	51 (7.2%)	
Total	680 (96.0%)	28 (4.0%)	708 (100%)	

*Sens* Sensitivity, *Spec* specificity, *PPP* positive predictive power, *NPP* negative predictive power, *Overall* overall correct classification rate,  $\kappa$  kappa statistic (agreement corrected for chance)



cantly related to ADS-IV ADHD inattentive/combined types (no vs yes),  $\chi^2(1)=160.45$ ,  $p<0.001$ . Diagnostic statistics (see Table 6) showed that the specificity (true negatives) was higher than the sensitivity (true positives) and the negative predictive power was higher than the positive predictive power for each IOWA scale. These results indicate that the selected cutoffs are more accurate at ruling out children who do not have IO, OD, HY or IN problems than they are at correctly identifying children who do have such problems. This is a good characteristic of a screening device.

**Teacher Ratings** The 2×2 chi-square analyses showed: (a) IOWA IO (no vs yes) was significantly related to ADS-IV ADHD-any subtype (no vs yes),  $\chi^2(1)=471.37$ ,  $p<0.001$ ; (b) IOWA OD (no vs yes) was significantly related to ADS-IV ODD (no vs yes),  $\chi^2(1)=450.81$ ,  $p<0.001$ ; (c) IOWA HY (no vs yes) was significantly related to ADS-IV ADHD hyperactive/impulsive/combined types (no vs yes),  $\chi^2(1)=396.91$ ,  $p<0.001$ ; (d) IOWA IN (no vs yes) was significantly related to ADS-IV ADHD inattentive/combined types (no vs yes),  $\chi^2(1)=529.65$ ,  $p<0.001$ . Diagnostic statistics (see Table 7) showed that the specificity (true negatives) was higher than the sensitivity (true positives) and the negative predictive power was higher than the positive predictive power for each IOWA scale except the OD scale which had high sensitivity and specificity. These results indicate that the selected cutoffs are more accurate at ruling out children who do not have IO, OD, HY or IN problems than they are at correctly identifying children who

do have such problems. This is a good characteristic of a screening device.

**Discussion**

This study examined the factor structure and psychometric properties of mother and teacher ratings on the IOWA Conners Rating Scale. Based on the a priori model of the IOWA and on previous research and theory, a two factor model was tested, which consisted of an inattentive-impulsive-overactive (IO) factor and an oppositional-defiant (OD) factor, and a three factor model was tested in which the IO factor was further divided into a hyperactive/impulsive (HY) factor and an inattentive (IN) factor. Confirmatory factor analysis was used to test the hypothesized models and found that the three factor model provided the best absolute fit to the observed data and showed a significantly better fit relative to the two-factor model. These results are consistent with current diagnostic formulations of ADHD and ODD and with a large body of research suggesting that oppositional defiant, hyperactive/impulsive, and inattentive behaviors are independent but correlated behaviors (Hinshaw 1987; Waschbusch 2002).

Confirmatory factor analyses also demonstrated that the IOWA as completed by mothers and teachers are functionally equivalent. This has commonly been assumed to be true, but this assumption has rarely been tested. For both informants the three factor model was the most well supported, and the items appear to relate to the factors in

**Table 7** Number (percent) of children in IOWA groups and ADS-IV groups as rated by teachers

	ADS-IV ADHD-any subtype		Total	Diagnostic statistics
	No	Yes		
IOWA IO				
No	1289 (87.0%)	42 (2.8%)	1331 (89.8%)	Sens=0.66; Spec=0.95 PPP=0.55; NPP=0.97
Yes	68 (4.6%)	83 (5.6%)	151 (10.2%)	
Total	1357 (91.6%)	125 (8.4%)	1482 (100%)	Overall=0.9; $\kappa$ =0.56
IOWA OD	ADS-IV ODD			
No	1343 (90.6%)	1 (0.1%)	1344 (90.7%)	Sens=0.98; Spec=0.94 PPP=0.33; NPP=0.99
Yes	92 (6.2%)	46 (3.1%)	138 (9.3%)	
Total	1435 (96.8%)	47 (3.2%)	1482 (100%)	Overall=0.9; $\kappa$ =0.47
IOWA HY	ADS-IV ADHD-HY and combined			
No	1316 (88.8%)	8 (0.5%)	1324 (89.3%)	Sens=0.87; Spec=0.93 PPP=0.34; NPP=0.99
Yes	104 (7.0%)	54 (3.6%)	158 (10.7%)	
Total	1420 (95.8%)	62 (4.2%)	1482 (100%)	Overall=0.9; $\kappa$ =0.46
IOWA IN	ADS-IV ADHD-IN and combined			
No	1314 (88.7%)	23 (1.6%)	1337 (90.2%)	Sens=0.77; Spec=0.94 PPP=0.52; NPP=0.98
Yes	70 (47.7%)	75 (5.1%)	145 (9.8%)	
Total	1384 (93.4%)	98 (6.6%)	1482 (100%)	Overall=0.9; $\kappa$ =0.58

Sens Sensitivity, Spec specificity, PPP positive predictive power, NPP negative predictive power, Overall overall correct classification rate,  $\kappa$  kappa statistic (agreement corrected for chance)

an equivalent way. Further, the IOWA scales were highly reliable for both mothers and teachers.

Analyses of grade and gender differences showed that scores on each of the IOWA scales showed small to moderate differences between boys and girls but showed small to insignificant differences between grades. This pattern was found for both mother ratings and teacher ratings. It is not surprising that boys had higher scores on the IOWA than did girls as this is consistent with a large body of other research (e.g., Gaub and Carlson 1997; Keenan et al. 1999; Waschbusch et al. 2006). The finding that neither IO nor OD ratings differed as a function of grade, and that the grade differences for HY and IN were very small, is consistent with other studies that suggest IO and OD behaviors tend to be relatively stable over the course of elementary school years (Cohen et al. 1993; DuPaul et al. 1997; Pelham et al. 1992). Alternatively, this response may indicate that the informants adjust their thresholds for using the Likert response scales as a function of the age of the child that they are rating (e.g., parents and teachers may use different criteria for determining what constitutes “pretty much” fidgeting for first versus fourth graders).

Based on these findings, cutoffs scores for identifying children with elevated IOWA scores were proposed, with separate cutoffs computed for boys and girls. Analyses to evaluate the proposed cutoffs showed they performed reasonably well in that they were significantly related to diagnoses based on parent and teacher ratings of the full range of DSM-IV symptoms for ADHD and ODD, including evidence of symptom related impairment. However, the proposed cutoffs come from a largely Caucasian population of students who resided in non-urban settings. The generalizability of these cutoffs to other racial groups and/or geographic regions is uncertain.

Overall, this study makes a number of contributions that have considerable practical significance. First, the results show that the IOWA Conners has good psychometric properties when used by mothers and teachers. This is the first study to evaluate the psychometric properties of the instrument when used by mothers. Second, contrary to common use, this study indicated that the items on the IOWA IO scale are better represented by two separate scales, one measuring inattentive behavior and a second measuring hyperactive/impulsive behavior. This method of scoring the IOWA is consistent with recent factor analytic work and with the current diagnostic formulation of ADHD. Third, the proposed cutoff scores hold considerable promise for using the IOWA Conners as an initial screening measure for identifying children with externalizing behavior problems, although further research on their validity is certainly needed.

There were several limitations of this study. First, although racial information was not collected on individual

children, the sample consisted largely of teachers and mothers of middle class, rural and Caucasian children. This reflected the communities the schools served, but *generalization to other populations should be done cautiously*. Second, it is possible that children of mothers who chose not to participate differ systematically from children of mothers who did participate in the study. For instance, it may be that some mothers elected not to participate because their children have especially high levels of disruptive behavior which they did not wish to call attention to. Alternatively, it may be that mothers elected not to participate because they did not feel the rating scale was relevant to their child. As described earlier, comparison of students with and without complete mother ratings suggested this is not the case as they did not differ on demographic measures (age, sex) or on teacher ratings of behavior. Even so, previous research suggests that mothers who participate in school-based research may differ from those who do not (Anderman et al. 1995), but how these findings apply to the present study is unclear. Third, it may be that schools that did not choose to participate differed systematically from those who did participate. While the schools included in the study appeared to be similar to other schools in the district, data to evaluate this empirically are not available. Fourth, although the analyses provided good support for distinguishing OD, HY and IN behaviors, both within and across informants, as well as strong indication of good internal consistency and retest reliability, absolute model fit was modest in some cases (see Table 2). These findings were likely due in part to large sample sizes (yielding high power to identify misfit), but the modest fit may also reflect imperfections in the construction of IOWA. For example, the IOWA was designed to yield an overall inattentive/impulsive/overactive (IO) scale even though more recent theory and research suggests inattentive behaviors should be distinguished from hyperactive/impulsive behaviors (e.g., Lahey et al. 1988; Milich et al. 2001; Pillow et al. 1998). Although the results supported distinguishing inattentive and hyperactive/impulsive behaviors as indexed by the IOWA Conners, the ability of this study to do so was limited by the low number of items used to measure these behaviors (3 HY and 2 IN). Moreover, the phrasing of some the OD items (e.g., acts ‘smart’, quarrelsome) may not be interpreted in the same way across informants. Nonetheless, the practical advantages of a free and brief screening measure for disruptive behavior outweigh many of these concerns.

In sum, these results indicate that the IOWA Conners is a psychometrically sound measure of externalizing behavior when completed by either mothers or teachers. Consistent with modern conceptualizations of the disruptive behavior disorders, a model that distinguished inattentive, hyperactive/impulsive, and oppositional behavior best represented

the observed data. Nonetheless, there was modest support for current practice of using IO and OD scales. Users should appreciate that this ‘traditional’ scoring method confounds two distinct dimensions of behavior. This may be an important limitation in some situations (e.g., in studies interested in identifying children with different subtypes of ADHD) but may not be important in other situations (e.g., in medication trials involving children who are already diagnosed with combined type ADHD; for screening purposes where the goal is to differentiate pure ADHD from ADHD with co-occurring conduct problems). More generally, the results of this study, in combination with the fact that the IOWA Conners is an exceptionally brief measure that is in the public domain, provide a strong basis for its continued use in both clinical and research settings.

Additional research further evaluating the IOWA would benefit the many users of the IOWA. Of particular import is research using longitudinal designs to examine age changes in the IOWA, as well as research examining the validity of the IOWA scales and of the proposed screening cutoffs. Alternatively, research aimed at developing an updated version of the IOWA also seems warranted. Much progress has been made in conceptualizing disruptive behavior in children in the more than 20 years since the IOWA first developed. There is now consensus that inattentive behaviors are conceptually and empirically distinct from overactive-impulsive behaviors, as discussed earlier, but this distinction was not part of the a priori model of the IOWA. While the results of this study show that the distinction between inattentive and overactive-impulsive behavior can be made using the current version of the IOWA, it is likely that a more clear distinction between the constructions could be found if the items comprising the IOWA were updated with this purpose in mind. This would greatly benefit researchers and clinicians seeking to conduct brief evaluations of externalizing behavior in children.

**Acknowledgement** This project was supported by grants from the Nova Scotia Health Research Foundation and the Social Sciences and Humanities Research Council (SSHRC). We would like to thank the many children, parents, teachers, and research assistants who helped make this project possible.

## References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213–232.
- Airaksinen, E. M., Michelsson, K., & Jokela, V. (2004). The occurrence of inattention, hyperactivity, impulsivity, and coexisting symptoms in a population study of 471 6–8 year old children based on the FTF (Five to Fifteen) questionnaire. *European Child and Adolescent Psychiatry*, *13*(Supplement 3), 23–30.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th text revision ed.)*. Washington, DC: American Psychiatric Association.
- Anderman, C., Cheadle, A., Curry, S., Diehr, P., Shultz, L., & Wagner, E. (1995). Selection bias related to parental consent in school-based survey research. *Evaluation Review*, *19*, 663–674.
- Atkins, M. S., Pelham, W. E., & Licht, M. H. (1988). The development and validation of objective classroom measures for the assessment of conduct and attention deficit disorders. In R. J. Prinz (Ed.), *Advances in behavioral assessment of children and families* (Vol. 4, pp. 3–33). Greenwich, CT: JAI.
- Atkins, M. S., Pelham, W. E., & Licht, M. H. (1989). The differential validity of teacher ratings of inattention/overactivity and aggression. *Journal of Abnormal Child Psychology*, *17* (4), 423–435.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unified theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94.
- Brent, D. A., Oquendo, M., Birmaher, B., Greenhill, L. L., Kolko, D., Stanley, B., et al. (2004). Familial transmission of mood disorders: Convergence and divergence with transmission of suicidal behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(10), 1259–1266.
- Burns, G. L., Walsh, J. A., Patterson, D. R., Holte, C. S., Sommers-Flanagan, R., & Parker, C. M. (1997). Internal validity of the disruptive behavior disorder symptoms: Implications from parent ratings for a dimensional approach to symptom validity. *Journal of Abnormal Child Psychology*, *25*(4), 307–319.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105* (3), 456–466.
- Cohen, P., Cohen, J., Kasen, S., Velez, C. N., Harkness, C., Johnson, J., et al. (1993). An epidemiological study of disorders in late childhood and adolescence: I. Age and gender specific prevalence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *34*(6), 851–867.
- Conners, C. K. (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry*, *126*, 884–888.
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K. E., & Ikeda, M. J. (1997). Teacher ratings of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Psychological Assessment*, *9*(4), 436–444.
- Elgar, F. J., Curtis, L. J., McGrath, P. J., Waschbusch, D. A., & Stewart, S. H. (2003). Antecedent-consequence conditions in maternal mood and child behavioural problems: A four-year cross-lagged study. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 362–374.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491.
- Frissell, K. C., McCarthy, D. M., D’Amico, E. J., Metrik, J., Ellingstad, T. P., & Brown, S. A. (2004). Impact of consent procedures on reported levels of adolescent alcohol use. *Psychology of Addictive Behaviors*, *18*(4), 307–315.
- Gadow, K. D., Sprafkin, J., Salisbury, H., Schneider, J., & Loney, J. (2004). Further validity evidence for the teacher version of the Child Symptom Inventory—4. *School Psychology Quarterly*, *19* (1), 50–71.
- Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(8), 1036–1045.
- Gottfredson, G. D., & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Educational and Psychological Consultation*, *12*(4), 313–344.

- Halperin, J. M., McKay, K. E., Grayson, R. H., & Newcorn, J. H. (2003). Reliability, validity, and preliminary normative data for the Children's Aggression Scale—Teacher version. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32(8), 965–971.
- Hart, E. L., Lahey, B. B., Loeber, R., & Hanson, K. S. (1994). Criterion validity of informants in the diagnosis of disruptive behavior disorders in children: A preliminary study. *Journal of Consulting and Clinical Psychology*, 62(2), 410–414.
- Hinshaw, S. P. (1987). On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Review*, 101(3), 443–463.
- Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling*, 5(4), 344–364.
- Johnston, C., & Pelham, W. E. (1986). Teacher ratings predict peer ratings of aggression at 3-year follow-up in boys with attention deficit disorder with hyperactivity. *Journal of Consulting and Clinical Psychology*, 54(4), 571–572.
- Keenan, K., Loeber, R., & Green, S. (1999). Conduct disorder in girls: A review of the literature. *Clinical Child and Family Psychology Review*, 2(1), 3–19.
- Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment*, 5(4), 395–399.
- Lahey, B. B., Miller, T. I., Gordon, R. A., & Riley, A. W. (1999). Developmental epidemiology of the disruptive behavior disorders. In H. C. Quay & A. E. Hogan (Eds.), *Handbook of disruptive behavior disorders* (pp. 23–48). New York: Kluwer.
- Lahey, B. B., Pelham, W. E., Schaugnecy, E. A., Atkins, M. S., Murphy, H. A., Hynd, G., et al. (1988). Dimensions and types of attention deficit disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 27, 330–335.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*, 1(4), 317–337.
- Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich & D. K. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 3, pp. 113–147). Greenwich, CT: JAI.
- McGrew, K. S., & Gilman, C. J. (1991). Measuring the perceived degree of parent empowerment in home-school relationships through a home-school survey. *Journal of Psychoeducational Assessment*, 9(5), 353–362.
- McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 477–505.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Milich, R., Balentine, A. C., & Lynam, D. R. (2001). ADHD combined type and ADHD predominately inattentive type are distinct and unrelated disorders. *Clinical Psychology: Science and Practice*, 8(4), 463–468.
- Milich, R., & Fitzgerald, G. (1985). Validation of inattention/overactivity and aggression ratings with classroom observations. *Journal of Consulting and Clinical Psychology*, 53(1), 139–140.
- Milich, R., & Landau, S. (1988). Teacher ratings of inattention/overactivity and aggression: Cross-validation with classroom observations. *Journal of Clinical Child Psychology*, 17(1), 92–97.
- Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript, University of California Los Angeles.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Author.
- NIMH-DISC Editorial Board. (1999). *The National Institute of Mental Health diagnostic interview schedule for children*. New York: Columbia University.
- Nolan, E. E., & Gadow, K. D. (1994). Relation between ratings and observations of stimulant drug response in hyperactive children. *Journal of Clinical Child Psychology*, 23(1), 78–90.
- Nova Scotia Department of Education. (2003). Statistical summary of Nova Scotia Schools (<http://stats-summary.ednet.ns.ca/index.shtml>).
- Nova Scotia Department of Finance. (2003). 2001 Census of Canada: Nova Scotia Perspective. Retrieved from <http://www.gov.ns.ca/finance/publish/>.
- Oosterlan, J., Sheres, A., & Sergeant, J. A. (2005). Which executive functioning deficits are associated with AD/HD, ODD/CD, and comorbid AD/HD plus ODD/CD? *Journal of Abnormal Child Psychology*, 33(1), 69–85.
- Pelham, W. E., Burrows-MacLean, L., Gnagy, E. M., Fabiano, G. A., Coles, E. K., Tresco, K. E., et al. (2005a). Transdermal methylphenidate, behavioral, and combined treatment for children with ADHD. *Experimental and Clinical Psychopharmacology*, 13(2), 111–126.
- Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005b). Evidence-based assessment for attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 449–476.
- 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(2), 210–218.
- Pelham, W. E., Milich, R., Murphy, D. A., & Murphy, H. A. (1989). Normative data on the IOWA Conners teacher rating scale. *Journal of Clinical Child Psychology*, 18, 259–262.
- Pillow, D. R., Pelham, W. E., Hoza, B., Molina, B. S. G., & Stultz, C. H. (1998). Confirmatory factor analyses examining attention deficit hyperactivity disorder symptoms and other childhood disruptive behaviors. *Journal of Abnormal Child Psychology*, 26(4), 293–309.
- Quay, H. C., & Hogan, A. E. (Eds.). (1999). *Handbook of disruptive behavior disorders*. New York: Kluwer.
- Remschmidt, H., Hoare, P., Etrick, C., Rothenberger, A., Santosh, P., Schmidt, M., et al. (2005). Symptom control in children and adolescents with attention-deficit/hyperactivity disorder on switching from immediate-release MPH to OROS (r) MPH—Results of a 3-week open-label study. *European Child and Adolescent Psychiatry*, 14(6), 297–304.
- Satorra, A., & Bentler, P. M. (1999). *A scaled difference chi-square test statistic for moment structure analysis*. Los Angeles: UCLA Statistics Series (paper # 260).
- Waschbusch, D. A. (2002). A meta-analytic examination of comorbid hyperactive/impulsive/inattention problems and conduct problems. *Psychological Bulletin*, 128, 118–150.
- Waschbusch, D. A., King, S., Gregus, A., & Northern Partners In Action for Children and Youth. (2007). Age of onset of ADHD in a sample of elementary school students. *Journal of Psychopathology and Behavioral Assessment*, 27, 9–16.
- Waschbusch, D. A., King, S., & Northern Partners In Action for Children and Youth. (2006). Should sex-specific norms be used to assess attention-deficit hyperactivity disorder (ADHD) or oppositional defiant disorder (ODD)? *Journal of Consulting and Clinical Psychology*, 74(1), 179–185.
- Waschbusch, D. A., Pelham, W. E., Massetti, G. M., & Northern Partners In Action for Children and Youth. (2005). The Behavior Education Support and Treatment (BEST) school intervention program: Pilot project data examining school-wide, targeted-

- school, and targeted-home approaches. *Journal of Attention Disorders*, 9(1), 313–322.
- Waschbusch, D. A., Sparkes, S. J., & Northern Region Partners In Action for Children and Youth. (2003). Rating scale assessment of attention-deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) symptoms: Is there a normal distribution and does it matter? *Journal of Psychoeducational Assessment*, 21(3), 261–281.
- Wolraich, M. L., Lambert, E. W., Bickman, L., Simmons, T., Doffing, M. A., & Worley, K. A. (2004). Assessing the impact of parent and teacher agreement on diagnosing attention-deficit/hyperactivity disorder. *Developmental and Behavioral Pediatrics*, 25(1), 41–47.
- Yu, C.-Y. (2003). *Evaluating cutoff criteria of model fit indices for latent variables with binary and continuous outcomes*. UCLA, Los Angeles.