Psychological Assessment

Parent–Teacher Agreement on ADHD Symptoms Across Development

Megan E. Narad, Annie A. Garner, James L. Peugh, Leanne Tamm, Tanya N. Antonini, Kathleen M. Kingery, John O. Simon, and Jeffery N. Epstein

Online First Publication, September 15, 2014. http://dx.doi.org/10.1037/a0037864

Psychological Assessment

1040-3590/14/$12.00 http://dx.doi.org/10.1037/a0037864

Parent–Teacher Agreement on ADHD Symptoms Across Development

Megan E. Narad, Annie A. Garner, James L. Peugh, Leanne Tamm, Tanya N. Antonini,
Kathleen M. Kingery, John O. Simon, and Jeffery N. Epstein
Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

Parent–teacher agreement on attention-deficit/hyperactivity disorder (ADHD) symptom ratings ranges from low to moderate. Most studies evaluating parent–teacher agreement have not assessed measurement invariance across raters. Hence, it is unclear whether discordance across raters is due to differing ADHD constructs across raters or other factors (e.g., subjective differences across raters). Additionally, the effect of development on parent–teacher agreement is relatively unknown. To address these limitations, the present study used parent and teacher ADHD ratings from a large ($N = 6{,}659$) developmentally diverse (ages 4–17) sample. Using exploratory structural equation modeling on half the sample, and then confirmatory factor analysis (CFA) on the other half of the sample, confirmed a 2-factor structure with significant cross-loadings for the 18 ADHD symptoms. CFA invariance analyses demonstrated that the 2-factor symptom structure was similar across raters and age groups. After confirming measurement invariance, the correlation between latent factors within and across raters was examined for each age group as well as across age groups. Parents reported greater severity of ADHD symptoms than did teachers, and both parents and teachers reported higher levels of hyperactivity/impulsivity in younger children than in older children and consistent levels of inattention across development. Finally, correlations between parent–teacher ratings of like factors were weak for inattention and moderate–strong for hyperactivity/impulsivity, and the magnitude of parent–teacher agreement did not vary across development. In conclusion, while parent and teacher ratings of ADHD behaviors are only weakly to moderately correlated, each reporter provides unique and valid clinical information as it relates to ADHD symptom presentation.

Keywords: parent–teacher agreement, interrater agreement, ADHD symptoms

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common psychiatric disorders of childhood, affecting an estimated 8% of school-aged children (Froehlich et al., 2007). ADHD is defined by developmentally inappropriate levels of inattention (IA) and/or hyperactivity/impulsivity (HI) that are impairing to the patient in more than one setting (American Psychiatric Association, 2000). Consensus clinical guidelines recommend that clinicians assess the pervasiveness of ADHD symptoms and impairments by collecting ratings of behavior in multiple environments, typically home and school (American Academy of Child and Adolescent Psychiatry, 2007; American Academy of Pediatrics, 2011). Moreover, the recent update of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5;* American Psychiatric Association, 2013) requires the presence of ADHD symptoms across multiple environments. Unfortunately, ratings from parents and teachers are often discrepant, which can impact a clinician's case conceptualization and differential diagnosis (Murray et al., 2007).

Low to moderate parent–teacher agreement on ADHD behavioral ratings is a common research finding (range of correlations = .09–.43; Antrop, Roeyers, Oosterlaan, & Van Oost, 2002; Mitsis, McKay, Schulz, Newcorn, & Halperin, 2000; Murray et al., 2007; Sollie, Larsson, & Morch, 2013; Willcutt et al., 2012; Wolraich et al., 2004). Whether parent–teacher agreement differs across the two ADHD symptom domains (i.e., IA and HI) is unclear. Some studies have reported higher parent–teacher agreement for IA symptoms (Antrop et al., 2002; Sibley et al., 2012; Wolraich et al., 2004), while others have reported higher parent–teacher agreement for HI symptoms (Mitsis et al., 2000; Sollie et al., 2013). Variability in methods across studies (i.e., sample size, age range, and congruency of rated environment) may contribute to the varying results across studies.

Megan E. Narad, Annie A. Garner, James L. Peugh, Leanne Tamm, Tanya N. Antonini, Kathleen M. Kingery, John O. Simon, and Jeffery N. Epstein, Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.

Tanya N. Antonini is now at the Section of Psychology, Department of Pediatrics, Baylor College of Medicine/Texas Children's Hospital, Houston, Texas.

Megan E. Narad and Annie A. Garner contributed equally to this article and share first authorship. Jeffery N. Epstein and the Cincinnati Children's Hospital Medical Center own the intellectual property related to the ADHD Web portal myADHDportal.com. Annie A. Garner was supported by funds from the Bureau of Health Professions (BHPr), Health Resources and Services Administration (HRSA), and the Development of Health and Human Services (DHHS) under Grant T32HT1002 National Research Service Award. The information or content and conclusions are those of the authors and should not be construed as the official position or policy of, nor should any endorsements be inferred by, the BHPr, HRSA, or DHHS of the U.S. government.

Correspondence concerning this article should be addressed to Megan E. Narad, Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229. E-mail: megan.narad@cchmc.org

There are several possible sources for disagreement between parent and teacher symptom ratings. For example, due to the subjective nature of rating scales, differences in ratings may reflect differences in rater perceptions of developmentally appropriate behavior, rather than true differences in symptom presentation across home and school settings (Amador-Campos, Forns-Santacana, Guardia-Olmos, & Pero-Cebollero, 2006). Discrepancies in ratings may also result from reporter bias, which could produce differing magnitudes of ratings across parents and teachers. For example, parents tend to report a higher magnitude of ADHD symptoms than teachers (Antrop et al., 2002; Hart, Lahey, Loeber, Applegate, & Frick, 1995; Murray et al., 2007; Papageorgiou, Kalyva, Dafoulis, & Vostanis, 2008). Potential reasons for higher ratings by parents than teachers include the impact of a less structured home setting compared to typical school settings, the greater amount of time and varying times of day that parents spend with their children compared to teachers, and so forth.

While systematic variability across parents and teachers may be one source of observed discrepancies between parent and teacher ratings, another reason for discordant ratings across parents and teachers is that the ADHD construct may differ across raters. Most studies examining inter-rater agreement have assumed invariance in the ADHD construct across raters and therefore have relied on simple correlations to examine the level of agreement between raters. Burns and colleagues (2008) suggested that concordance in parent and teacher ratings should be tested using latent variable multiple-group modeling with confirmatory factor analysis (CFA) so that the consistency and validity of ADHD symptom scores across raters is established before estimating agreement. In fact, the agreement among reporters cannot be estimated without first establishing that the instrument used to measure symptoms is in fact measuring the same construct across raters (Chen, Sousa, & West, 2005). Without first establishing that the ADHD construct is invariant across raters, it will remain unclear whether parent–teacher disagreement reflects true differences in symptom presentation or is an artifact of measurement variance across raters.

There are additional advantages to using latent variable modeling procedures, such as multiple-group CFA, to assess inter-rater agreement. In particular, the use of latent modeling of ADHD symptoms produces error-free dimensions, which allows correlational estimates to more clearly reflect variation across raters and not measurement error. Second, in addition to allowing error-free estimates of parent–teacher agreement across ADHD symptom domain ratings (e.g., parent and teacher agreement on symptoms of IA), this procedure also generates estimates for within-rater, cross-domain associations (e.g., parent-rated IA and parent-rated HI) and cross-rater, cross-domain associations (e.g., parent-rated IA and teacher-rated HI). This full complement of rater–domain association estimates allows estimation of convergent and divergent validity.

While latent variable multiple-group modeling has been used to examine inter-rater agreement on ADHD symptom rating scales (e.g., Burns et al., 2008; Burns, Walsh, Gomez, & Hafetz, 2006), it has typically been used to examine cross-rater agreement in congruent settings (i.e., mother and father ratings). Burns et al. (2013) did include teacher ratings in addition to mothers and fathers ratings in a study using Spanish and Thai adolescents and found strong agreement between mothers' and fathers' ratings of ADHD and oppositional defiant disorder (ODD) across both Thai (r = .70–.73) and Spanish (r = .89–.91) parents. However, parent–teacher agreement was weak in the Thai group (r = .09–.27) and moderate to strong in the Spanish group (r = .45–.62). These results suggest that parent–teacher agreement may vary according to culture or possibly according to instruments, since different scales were used across the Thai (i.e., Child and Adolescent Disruptive Behavior Inventory; Burns, Taylor, & Rusby, 2001a, 2001b) and Spanish (i.e., ADHD Rating Scale IV; DuPaul, Power, Anastopoulos, & Reid, 1998) samples. There is a clear need to obtain estimates of parent–teacher agreement using similar statistical methods with a U.S. sample. Moreover, given that the Vanderbilt ADHD Rating Scales (Wolraich, Feurer, Hannah, Baumgaertel, & Pinnock, 1998) is a frequently used ADHD diagnostic symptom checklist used in pediatric settings (American Academy of Pediatrics, 2011), and examining parent–teacher agreement on this measure is clinically relevant.

In addition to subjective bias and measurement issues, there exist other potential variables that may affect rater agreement (see De Los Reyes & Kazdin, 2005). Foremost of these issues is the child's age at the time of assessment. It is well known that symptom presentation changes across development, with decreasing levels of HI symptoms and fairly stable IA symptoms with increasing age (Harpin, 2005; Larsson, Larsson, & Lichtenstein, 2004; Sibley et al., 2012). However, the impact of this heterotypic continuity among ADHD symptom domains on parent–teacher agreement is unknown.

The existing literature on ratings of behaviors is rather mixed in regard to the effects of child's age on rater agreement (Achenbach, McConaughy, & Howell, 1987; Choudhury, Pimentel, & Kendall, 2003; Engel, Rodrigue, & Geffken, 1994; Grills & Ollendick, 2003; Jensen, Xenakis, Davis, & Degroot, 1988; Kolko & Kazdin, 1993; Verhulst, Althaus, & Berden, 1987). Murray and colleagues (2007) reported low to moderate parent–teacher agreement on ADHD ratings in preschool children (r = .24–.26), while Sibley and colleagues (2012) reported slightly higher levels of agreement on ADHD ratings (r = .35–.41) using an adolescent population. Together, these results suggest a possible trend toward higher parent–teacher agreement with increasing age. Given the developmental trend in symptom presentation whereby symptoms of HI are more prominent in preschool while symptoms of IA are more prominent in adolescence (Harpin, 2005; Larsson, Larsson, & Lichtenstein, 2004; Sibley et al., 2012), these findings also suggest that level of parent–teacher agreement may vary across symptom domain. Ultimately, these potential age- and ADHD symptom domain-specific trends need to be evaluated within a single developmentally diverse sample. To date, only one study (Sayal & Goodman, 2009) has examined the impact of age on parent–teacher agreement in a single sample of children aged 5–16. While the authors found that age did not moderate levels of parent–teacher agreement, their sample was divided into two broad age categories of younger (5–10 years) and older (11–16 years) children, which may have obscured any effects of age.

In the present study we attempt to address the limitations of previous work by using a large, developmentally diverse sample of clinic-referred children grouped into age categories based on school setting (preschool, elementary school, middle school, and high school). The goal of this study was to examine the level of parent–teacher agreement on the Vanderbilt ADHD Rating Scales across age groups. This was accomplished in the following steps:

1. In order to identify potential items with significant cross-loadings in the two-factor *DSM* model of ADHD, an exploratory structural equation model (ESEM) was conducted using half of the total sample (all age groups and raters).

2. The resulting model was then confirmed using CFA on the second half of the total sample.

3. To test for measurement invariance, CFA analysis within each age group was conducted in order to determine if the ADHD construct was equivalent at the factor (metric), threshold (scalar), and latent factor mean levels across parents and teachers.

4. Multiple-group (four age groups) CFA invariance analysis was performed to determine if the ADHD construct was equivalent at the factor (metric), threshold (scalar), and latent factor mean levels across age groups.

5. After confirming measurement invariance, the correlation between latent factors of IA and HI within and across raters was examined for each age grouping as well as across age groups.

We predicted that parent and teacher ratings of ADHD would have an equivalent two-factor structure (i.e., IA and HI factors) for all age groups, with equivalent factor loadings and thresholds (i.e., initial scale points, equivalent to intercepts for categorical data) across raters. Consistent with previous research (Antrop, Roeyers, Oosterlaan, & Van Oost, 2002; Murray et al., 2007), we expected parent ratings to be higher than teacher ratings, and that symptom presentation would be consistent with heterotypic continuity (i.e., higher ratings on the HI dimension among the younger age groups and consistent levels of IA across age groups; Harpin, 2005; Larsson, Larsson, Lichtenstein, 2004; Sibley et al., 2012). We expected to observe low to moderate levels of agreement between parent and teacher ratings on like symptom dimensions. Based on developmental trends in the literature (Murray et al., 2007; Sibley et al., 2012), we expected to find the highest levels of parent–teacher agreement among older adolescents.

Method

Participants

Participants were patients registered on an Internet-based ADHD web portal being used by 234 primary care and mental health professionals throughout the United States (including California, Georgia, Illinois, Kentucky, Massachusetts, Missouri, New York, Ohio, Texas, and Wisconsin) to collect parent and teacher rating scales as part of their ADHD diagnostic assessments with children. From October 2008 to June 2013, these 234 professionals registered 8,466 families with children between the ages of 4 and 17 on the ADHD web portal for ADHD assessments. All children had a completed parent rating form, and 6,659 (78.66%) of these children had at least one completed teacher rating form. For the purposes of this study, we selected only patients who had both parent and teacher symptom rating scales ($n = 6{,}659$). While those excluded from the sample as a result of a missing teacher rating had less severe parent-rated symptoms of IA, $t(8464) = 4.26$, $p < .001$, than those with teacher ratings, the magnitude of this difference was small (Cohen's $d = 0.09$). The two groups did not differ on parent-rated HI symptom severity, $t(8464) = 0.52$, $p = .61$.

Within the sample of children with parent and teacher ratings, 4,542 (68.2%) were male, the mean age was 9.34 ($\pm$3.03) years, and 586 (8.8%) were on medication at the time of evaluation. Children were grouped into four age categories: preschool (ages 4–5.99); elementary school (ages 6–10.99); middle school (ages 11–13.99), and high school (ages 14–17.99). See Table 1 for the breakdown of sample size by age group and sex.

Measures

Vanderbilt ADHD Rating Scales. The Vanderbilt, which comprises parent reports (VADPRS) and teacher reports (VADTRS; Wolraich et al., 1998), is based on the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text –rev.; *DSM–IV–TR;* American Psychiatric Association). Both rating scales include assessment of the 18 symptoms of ADHD as outlined in the *DSM–IV–TR*. Each ADHD symptom is rated on a 4-point scale ranging from 0 (*never*) to 3 (*very often*). Psychometric properties for the VADPRS reveal acceptable reliability (alpha range = .91–.94) and adequate concurrent criterion validity (correlation between the VADPRS and the Diagnostic Interview Schedule for Children, Version 4; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000 = .66–.69). With regard to construct validity, CFA of the VADPRS revealed a satisfactory fit (comparative fit index [CFI] = .91) between parent report and the two-factor (IA and HI) model of ADHD (Wolraich et al., 2003). Test–retest reliability correlations exceeded .80, and the scale has reasonable sensitivity and positive predictive power (Bard, Wolraich, Neas, Doffing, & Beck, 2013). Similarly for teachers, construct validity is acceptable (CFI > .90), convergent validity with the Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998) is high (Pearson's correlations > .72), and internal consistency and reliability is acceptable for the VADTRS (alpha = .89–.96; Wolraich, Bard, Neas, Doffing, & Beck, 2013; Wolraich et al., 2003). While studies examining the inter-rater agreement within settings are not available for the VADRS, inter-rater agreement is quite high for similar ADHD symptoms rating scales (i.e., ADHD Rating Scale; Burns et al., 2013).

For this study, where multiple teachers completed the VADTRS on the same patient, a single teacher rating was selected based on the subject area taught by the rater. A single teacher was used in order to have comparable ratings across age groupings (i.e., older age groups are more likely to have multiple ratings compared to

Table 1
Sample Sizes for Parent and Teacher Reports by Age Group

Age group	Females	Males	Total
Preschool (4–5.99 years)	160	468	628
Elementary school (6–10.99 years)	1,387	2,955	4,342
Middle school (11–13.99 years)	346	672	1,018
High school (14–17.99 years)	223	448	671
Total *N*	2,116	4,543	6,659

younger age groups). In selecting a single rating, we preferentially selected a single teacher rating from multiple teacher ratings based on the following order of subject areas: homeroom ($n = 2{,}152$), special education ($n = 396$), language arts ($n = 1{,}615$), math ($n = 518$), social studies/history ($n = 113$), science ($n = 120$), and other (health, computer, art, after school program, music, physical education, teacher's assistant and tutor; $n = 1{,}745$). The rationale for this ordering of teachers was as follows: First, academic subjects where children spend most of their day (i.e., homeroom and special education) were prioritized. Next, language arts and math subjects were selected, as these academic subjects are taught across the entire age range of participants in this study (i.e., 4–17 years old). Since there were more language arts ratings than math ratings, language arts was prioritized over math. Social studies and science occurred at similar frequencies and were therefore selected in an arbitrary order. The remaining subjects, which occurred very infrequently, were clustered into a single *other* category and selected only if data from one of the previously mentioned subject areas was not available. Only one parent was allowed to complete the VADPRS; therefore each child had one parent and one teacher rating. Information regarding the relationship of the parent rater to the child (e.g., mother, father, legal guardian) was not obtained.

Procedures

This study is a retrospective data analysis using data gathered from the ADHD web portal software. The ADHD web portal is a platform whereby parents, teachers, and health care providers all mutually input information about the patient, after which information is scored, interpreted, and formatted in a report style that is helpful to the health care provider in his/her assessment and treatment of patients with ADHD (Epstein, Langberg, Lichtenstein, Kolb, & Simon, 2013). Health care providers initiate an evaluation by registering families on the portal to facilitate the collection of parent and teacher rating scales. After the health care provider registers a family for an assessment on the web portal, the parent receives an e-mail message from the health care provider inviting him/her to complete an online version of the VADPRS. The parent is then able to log on to the portal and invite teachers to complete and online version of the VADTRS. All of the study procedures were approved by the Institutional Review Board.

Analyses

Mplus Version 7.11 (Muthén & Muthén, 1998–2012) was used for all statistical analyses. VADTRS items (0–3 scale) were treated as categorical in all analyses, and robust weighted least squares estimation was used.

Establishing a baseline model. The sample was stratified by age group and then was randomly split in half. In order to maximize power, parent and teacher ratings were merged, and the lack of independence across parent- and teacher- ratings was handled using the Type = Cluster command in Mplus. With one-half of the sample, an ESEM was conducted with the number of factors being restricted to two (IA and HI based on the well-established two-factor symptom structure of ADHD symptoms; Willcutt et al., 2012). Of note, items with significant cross-loadings (both $p < .001$ and loadings ≥ 0.10) across the two ADHD factors were identified so that cross-loadings could be modeled and confirmed via CFA using the second half of the sample. CFA model fit was assessed by inspecting fit indices including CFI and root-mean-square error of approximation (RMSEA) with good fit defined as CFI ≥ 0.95 and RMSEA ≤ 0.05, and acceptable fit defined as CFI ≥ 0.91 and RMSEA ≤ 0.08 (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004).

Model invariance across raters. We tested the invariance of the best fitting baseline model across parents and teacher raters simultaneously with CFA. Separate models were estimated for each of the four age groups. An unconstrained (configural) model was estimated that had the factor loadings and thresholds free across raters, scale factors fixed at one in both parent and teacher sources, and factor means fixed at zero in both sources. The metric of the factor was set by fixing the first factor loading to one and allowing factor variances to be free across raters. The equivalence of factor loadings was tested using specifications for metric invariance as outlined in Millsap (2011), where the factor loadings are constrained to be equal across raters, among other specifications (see Mplus Version 7.1 Language Addendum; Muthén & Muthén, 1998–2012b). The metric invariance model was compared to the configural model to determine whether the strength of the relationship between items and their latent factors was the same across raters. Typically, chi-square likelihood tests (e.g., the DIFFTEST command in Mplus) are used to compare nested (i.e., unconstrained vs. constrained) models when using categorical data; a significant difference test is considered evidence against invariance. However, this method is extremely sensitive to large sample sizes. Therefore, we instead used a method suggested by Little (2013) that is commonly used by others in the field (e.g., Burns et al.,, 2006). A decrement in model fit, defined as a decrease in CFI $\leq .01$ and/or an increase in RMSEA of $\geq .015$, from the configural model to the constrained model is evidence against invariance across raters.

Once metric invariance across raters was established, scalar invariance was tested using specifications outlined in Millsap (2011) where factor loadings and thresholds are constrained to be equal across sources. The metric invariance model was the new baseline model against which the scalar invariance model was tested. Again, changes in fit statistics were used to assess the equivalence of factor loadings and thresholds across parents and teachers where a decreased model fit in the scalar model would suggest rater differences at the threshold level. Finally, latent mean differences between parents and teachers within age groups were tested using the Model Constraint command in Mplus.

Model equivalence across age groups. We first tested for invariance of the two-factor, cross-loading model across the four age groups. This required a group factor for each age group (1 = Preschool, 2 = Elementary, 3 = Middle, and 4 = High school). The baseline model (configural model) held parent and teacher like-item factor loadings and thresholds to equivalence *within* age groups but allowed these parameters to be free *across* age groups. Metric invariance (factor loadings) across groups was tested by holding like-item factor loadings between parent and teachers and across age groups (additional specifications for metric invariance modeling were followed as outlined in Muthén & Muthén, 1998–2012). If support was found for the metric invariance model, this model became the baseline against which a scalar invariance model was compared. Scalar invariance across age groups was tested by holding like-item factor loadings and like-item thresholds equivalent between parent and teachers and across age groups (additional specifications for metric invariance modeling were followed as outlined in Muthén & Muthén, 1998–2012). Models were compared by evaluating change in model fit indices where a decrease CFI $\geq .01$ and an increase in RMSEA of $\geq .015$ from the baseline model to the constrained model would be indicative of a meaningful change.

In the event that scalar invariance was established, we could also test for latent mean invariance. Invariance was tested by specifying a model in which latent means across raters and age groups were free (unconstrained model), except for the latent means of the preschool group, which were set to zero for estimation purposes. This model was tested against a model in which the like-factor latent means within rater were set to equality across age groups (constrained model). Models were compared using change in model fit indices as previously described. If model fit worsened in the constrained model, then model fit indices were inspected to identify which latent mean should be freed until the difference between the constrained and partially constrained model were no longer meaningful according to change in fit criteria.

Parent–teacher agreement, discriminant validity, and source effects within age groups. As indicated by Little (2013), the model that establishes invariance at the strongest level was then used to estimate the relationships within- and between-raters and within- and between-ADHD symptom domains across each age group. Since Mplus estimates covariance rather than factor correlations by default, phantom factors for the IA and HI dimensions were created (see Little, 2013). Using these phantom factors, parent–teacher agreement on like symptoms was assessed (convergent validity). To determine discriminant validity, factors (IA and HI) were required to correlate less than .80–.85 within each reporter (Burns et al., 2013). Source effects were estimated by comparing the different-factor, same-source correlations (e.g., IA parent and HI parent ratings) to different-factor, different-source correlations (e.g., parent ratings of IA with teacher ratings of HI) separately for each age group. The extent to which the former is larger than the latter indicates the magnitude of the source effects.

Discriminant validity *between raters* was assessed within each age group by testing whether parent–teacher agreement on like factors (e.g., correlation between parent and teacher ratings of IA) was significantly larger than correlations across sources on different factors (e.g., correlation between parent IA and teacher HI). Significance tests for discriminant validity between sources and source effects were tested via the Model Constraint command in Mplus.

Parent–teacher agreement across age groups. In order to test whether the parent and teacher agreement estimates varied as a function of age groupings, correlations across age groups were compared using the multiple group (age) scalar invariance model. This was done by conducting a series of pairwise significance tests comparing the difference between convergent phantom correlations across the four age groups (12 pairwise comparisons total) using the Model Constraint command in Mplus.

Results

Establishing a Baseline Model

ESEM analyses with half of the sample revealed that three IA items—difficulty keeping attention (cross-loading = 0.19, $p < .001$; primary loading = .80, $p < .001$), not listening (cross-loading = 0.34, $p < .001$; primary loading = .55, $p < .001$), and easily distracted (cross-loading = 0.34, $p < .001$; primary loading = .61, $p < .001$)—loaded significantly on the HI latent factor, and two HI items—fidgets (cross-loading = 0.18, $p < .001$; primary loading = .70, $p < .001$) and leaves seat (cross-loading = 0.14, $p < .001$; primary loading = .80, $p < .001$)—loaded significantly on the IA latent factor. Using the second half of the sample, this two-factor, cross-loading model was validated using CFA (CFI = 0.97 and RMSEA = 0.08).

Model Equivalence Across Raters

When the parent and teacher two-factor models with cross-loadings were estimated together, model fit as indicated by CFI was acceptable for all age groups; however, RMSEA fit indices suggested that model fit was acceptable only for the middle school and high school groups (see Table 2). After constraining factor loadings and thresholds to be equal across parents and teachers (i.e., scalar invariance), there was no degradation of model fit, as indicated by trivial changes in CFI and RMSEA, suggesting that the parameter estimates are equivalent across parents and teacher ratings of ADHD (see Table 2).

Significant differences in latent means across parents and teachers were found for IA and HI dimensions. Parent ratings were significantly greater than teacher ratings within all age groups (range of the mean difference for IA = 0.10–0.86, $p < .001$; range of the mean difference for HI = 0.21–0.42, $p < .001$).

Model Equivalence Across Age Groups

When all age groups were modeled in a single configural model, the fit of the model was acceptable (see Table 3). Constraining like loadings to equality across age groups improved model fit in

Table 2
Model Fit Indices for Rater Agreement for All Age Groups Separately

Age group	Configural				Metric invariance				Scalar invariance			
	χ^2	*df*	CFI	RMSEA	χ^2	*df*	CFI	RMSEA	χ^2	*df*	CFI	RMSEA
Preschool	1,666.876	258	0.957	0.093	16,705.74	279	0.958	0.089	1,999.882	313	0.949	0.093
Elementary school	8,910.018	258	0.966	0.088	8,884.15	279	0.966	0.084	10,421.896	313	0.960	0.086
Middle school	1,795.485	258	0.979	0.077	1,944.517	279	0.977	0.077	2,117.910	313	0.975	0.075
High school	1,156.004	258	0.978	0.072	1,216.115	279	0.977	0.071	1,348.839	313	0.974	0.070

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation.

Table 3
Equivalence Across Age Groups

Model	χ^2	*df*	CFI	RMSEA
Configural	24,662.431	2620	0.931	0.071
Metric invariance	23,159.257	2629	0.936	0.068
Scalar invariance	23,832.119	2731	0.934	0.068

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation.

comparison to the configural model, as indicated by an increase in CFI and a reduction in RMSEA. This model became the model against which scalar invariance (constraining like thresholds) was tested. Constraining like thresholds across age groups (scalar invariance) did not result in significant model deterioration, as evidenced by minimal changes in CFI (reduction in CFI = .002) and no change in RMSEA. These analyses suggest that the two-factor model with cross-loadings fit equally well across age groups.

Latent mean invariance analyses suggested that the groups were not invariant at the latent mean level across age groups (see Table 4). Freeing latent means for parent and teacher ratings of HI while leaving IA latent means equal across age groups resulted in a model that was not statistically different from the unconstrained model. Inspection of latent factor means for HI suggests that parent and teacher ratings of HI are higher for younger children than for older children (preschool > elementary school > middle school = high school). Both parents and teachers report consistent levels of IA across age groups.

Parent–Teacher Agreement, Discriminant Validity, and Source Effects Within Age Groups

Using the scalar invariance models across parents and teachers within each age group, the correlations between phantom IA and HI factors based on parent and teacher ratings for each age group separately are presented in Table 5. Parent–teacher agreement on like-symptom factors (i.e., convergent validity) was weak for the IA factor (mean r = .23; r range = .18–.32) and moderate for the HI factor (mean r = .38; r range = .33–.44). Discriminant validity within sources was supported, as none of the correlations between IA and HI were greater than .80 for parents (mean r = .40; r range = .31–.56) or teachers (mean r = .48; r range = .42–.51). Same-source, different-factor correlations were significantly stronger than different-source, different-factor correlations (ps < .001;

Table 4
Latent Factor Mean Differences Across Age Groups

Model	χ^2	*df*	CFI	RMSEA
Unconstrained	16,528.536	2725	0.957	0.055
Fully constrained	27,570.890	2733	0.923	0.074
Partially constrained-teacher HI-free	21,507.710	2731	0.941	0.064
Partially constrained-teacher and parent HI-free	18,677.291	2729	0.950	0.059

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; HI = hyperactivity/impulsivity.

Table 5
Rater by Symptom Dimension Phantom Factor Correlations

Variable	Parent IA	Parent HI	Teacher IA	Teacher HI
		Preschool		
Parent				
IA				
HI	$.564^{***}$			
Teacher				
IA	$.210^{***}$	.015		
HI	.056	$.333^{***a}$	$.486^{***}$	
		Elementary school		
Parent				
IA				
HI				
Teacher	$.372^{***}$			
IA				
HI	$.184^{***b}$	.008		
IA	–.049	$.443^{**a,c}$	$.423^{***}$	
		Middle school		
Parent				
IA				
HI				
Teacher	$.347^{**}$			
IA				
HI	$.218^{***}$	$.128^{***}$		
IA	–.012	$.392^{***}$	$.494^{***}$	
		High school		
Parent				
IA				
HI				
Teacher	$.305^{***}$			
IA				
HI	$.320^{***b}$	$.131^{**}$		
IA	.067	$.354^{***c}$	$.507^{***}$	

Note. Alphabetical superscripts indicate that the correlations are significantly different from each other at $p < .01$. IA = inattention; HI = hyperactivity/impulsivity.
** $p < .01$. *** $p < .001$.

results are not presented here but are available from the author upon request), indicating a significant portion of the variance in ratings (16% of variance in parent ratings and 23% of variance in teacher ratings) could be attributed to the source of the rating. Same-factor, different-source correlations were significantly larger than the different-factor, different-source correlations, indicating support for discriminant validity between sources (ps < .01; results are not presented here but are available from the author upon request).

Parent–Teacher Agreement Across Age Groups

Differences in the strength of agreement between parent and teacher ratings of like factors were tested in a series of pairwise comparisons. Only three comparisons were statistically significant (indicated by superscripts in Table 5). Specifically, the parent–teacher correlations for HI were statistically stronger in the elementary school group (r = .44, p = .004) than in the preschool group (r = .33, p = .03) and high school group (r = .32, p = .001). However, the correlation between parent and teacher ratings of IA was statistically stronger in the high school group (r = .32) than in the elementary school group (r = .18, p = .001).

Discussion

A two-factor model of ADHD with some item cross-loadings across factors fit the parent and teacher ratings well across raters and age groups. The two-factor model with cross-loadings was consistent for both parents and teachers, suggesting that this factor structure for ADHD symptoms is consistent across home and school settings. Additionally, this two-factor model held true across age groups, suggesting that the conceptualization of ADHD as consisting of separate IA and HI symptom domains does not need to be modified based on age for children 4–17 years of age. Analyses comparing latent factor means suggest that parent ratings of children's behavior are statistically higher than teacher ratings of children. Examination of latent means across age groups revealed that while parents reported higher levels of symptoms than do teachers, both parents and teachers reported higher levels of symptoms of HI in younger children than in older children. However, both parents and teachers reported consistent levels of IA across development. Because parent and teacher ratings and ratings across age groups were invariant at the threshold and factor loading level, these differences in ratings are not due to differences in parameters at the factor-loading or threshold level across raters and are instead the result of systematic variability across raters and/or settings (i.e., environment demands, subjectivity of parent vs. teacher ratings, differences in symptom presentation across settings). Finally, parent–teacher agreement was greater for symptoms of HI than IA, and this pattern of agreement was largely the same across age groups.

Not only was the *DSM* two-factor model supported by our analyses, but ESEM analyses identified items that had cross-loadings on a secondary factor. The number of items that loaded on both factors in the present study (difficulty keeping attention, difficulty listening, easily distracted, fidgets with hands and feet, and leaves seat) is not unexpected, given existing knowledge about the ADHD symptom domains. Indeed, the IA and HI dimensions are highly correlated (r range = .56–.85; Conners, Sitarenios, Parker, & Epstein, 1998; DuPaul et al., 1997; Molina, Smith, & Pelham, 2001; Wolraich et al., 2003). Further, other studies have shown similar patterns of cross-loadings (e.g., Burns et al., 2013). There is little doubt that the ADHD symptom domains are highly inter-related. Therefore, these cross-loadings do not necessarily suggest a specificity problem with the VADRS, especially since the magnitude of the cross-loadings on the secondary factor was weaker than the magnitude of the items' loadings on the primary factor. It is possible that the wording of these items leaves too much room for interpretation, and clarifying these items may improve content validity. Interestingly, recent changes in the *DSM–5* include the provision of example behaviors for each symptom, thereby addressing this ambiguity. Future studies examining the psychometric properties of ADHD symptom rating scales using the newly modified symptom descriptions may be able to address this empirical question.

By establishing measurement invariance to the scalar level (equivalence at the threshold level) between raters and across age groups, our study is able to attribute these differences to true rater differences rather than measurement issues or construct differences. Our finding that parents rated children as having greater magnitudes of ADHD symptoms than did teachers is consistent with previous research (Antrop et al., 2002; Hart et al., 1995; Murray et al., 2007; Papageorgiou et al., 2008). Similarly, results reflect the well-known heterotypic continuity among ADHD symptom domains across development with decreasing levels of HI with increased age, and stable levels of IA symptoms (Harpin, 2005; Larsson, Larsson, & Lichtenstein, 2004; Sibley et al., 2012). These differences across raters and age groups are valid and clinically relevant differences, lending support to (a) American Academy of Pediatrics guidelines suggesting that information be gathered from multiple sources and (b) *DSM* diagnostic criteria requiring impairment in multiple settings, as well as recent (c) changes in *DSM–5* allowing for fewer symptoms of HI for older adolescents and young adults (American Academy of Child and Adolescent Psychiatry, 2007; American Academy of Pediatrics, 2011; American Psychiatric Association, 2000, 2013).

Strength of agreement across parents and teachers was stronger for symptoms of HI (r range = .33–.44) than IA (r range = .18–.32), suggesting that convergent validity between sources differs across ADHD symptom domains. While modest convergent validity for HI and weak to moderate convergent validity of IA symptom ratings may raise concerns regarding the psychometric properties of the Vanderbilt, similar findings regarding parent–teacher agreement on ADHD symptom dimension have been reported using a different rating scale (Child and Adolescent Disruptive Behavior Inventory; Burns, et al., 2013). Thus, problems with convergent validity across parents and teachers are more likely an indication of the role of environment and expectations than problems related to psychometric measurement. It should be noted that the convergent correlations of IA symptoms for the elementary school group in the present study is lower than those for the Spanish sample (which used a similar rating scale) presented by Burns and colleagues (2013); however, this may be related to sampling procedures. Whereas the present study used a sample composed of patients presenting for ADHD evaluation, Burns and colleagues used a normative sample and hence had a broader range of symptom presentation. A more symptomatic sample results in more opportunity for between-sources variability (i.e., since parents and/or teachers are likely to rate children high, there is more room for discrepancies between the two sources), thereby lowering the magnitude of correlations between raters.

There are number of possible reasons for higher parent–teacher agreement for HI symptoms. First, externalizing behaviors, such as symptoms of hyperactivity and impulsivity, are more observable (particularly in young children) than are symptoms of IA and thus may be easier for parents and teachers to identify (Achenbach, McConaughy, & Howell, 1987; Evans, Allen, Moore, & Strauss, 2005). Additionally, these symptoms may be more disruptive to home and classroom environments, thereby making it more likely that they would be identified and focused on by parents and teachers. In addition, because of the disruptive nature of HI symptoms in the classroom, teachers are more likely to discuss these symptoms with parents, which may cause HI symptoms to be rated more similarly across raters (Evans et al., 2005). Yet another explanation is that HI symptoms are less likely to be impacted by the home versus the classroom environment. That is, teachers are likely to see IA symptoms in the classroom because they observe children during tasks that require sustained attention and concentration. However, parents have fewer opportunities to observe their children in comparable situations, as homework may not be monitored by parents as closely as classwork is monitored by teachers.

Cross-rater agreement of non-like symptoms was rather weak (absolute r range = .007–.132) and significantly weaker than same-factor, different-source correlations (e.g., parent- and teacher-rated IA). This, in combination with the fact that neither of the correlations between IA and HI factors within sources was greater than .58, suggests that there is good discriminant validity between the IA and HI factors across and within sources. These findings provide further support for the overall validity of using two separate ADHD symptom domains to summarize and define ADHD symptom presentation in children and adolescents (Burns et al., 2013).

Also important for understanding the parent and teacher agreement on ratings of ADHD symptoms is understanding how much of the variance in ratings can be accounted for by factors related to the perceptions of the respondent (source effects). Estimates of non-like symptoms within reporter (e.g., parent IA and parent HI) tended to be larger (r range = .31–.56) than estimates of cross-rater agreement of non-like symptoms (e.g., parent IA and teacher HI; absolute r range = .01–.13). Based on these results, we conclude that, on average, between 16% and 23% of the variance in ratings can be attributed to source effects. In general, ratings of symptoms made by individuals who observe the child in congruent settings have a higher level of agreement than symptom ratings made by those who observe the child in incongruous settings (Achenbach et al., 1987). Moreover, these findings are consistent with those reported by Gomez et al. (2003) in their multi-trait, multi-source investigation of parent and teacher ratings of ADHD, where they found that source effects were equal to or exceeded parent–teacher agreement on like symptoms (e.g., trait effects). It has been argued that these findings are not surprising, given that rating scales measure adult perceptions of behavior and that child behavior is highly contingent on the environment (DuPaul, 2003). In contrast, others have argued that large source effects reflect problems with ADHD rating scales (see Burns, Gomez, Walsh, & de Moura, 2003, for a discussion). Unfortunately, although our findings support the notion that source effects play a significant role in parent and teacher ratings of ADHD, our findings cannot shed light on this particular debate. However, it seems as though both arguments have their merits.

Finally, results indicate that, in general, parent–teacher agreement on ratings of ADHD symptoms are not moderated by the age of the children, since the majority of comparisons were nonsignificant. The only significant differences observed in our sample involved the elementary school group. Interestingly, whether the elementary school group had higher agreement or lower agreement was dependent on the symptom domain. With regard to symptoms of IA, parent and teacher agreement was weak and statistically lower than in the high school group, where agreement was moderate in strength. In contrast, agreement on HI symptoms was statistically stronger in the elementary school group than in the preschool group and high school group. It should be noted that, though statistically significant, these correlations descriptively fall within the moderate range across all age groups.

While this study has a number of strengths, including the use of a large, geographically and developmentally diverse clinical sample presenting for ADHD assessment and a sophisticated statistical approach, it has noted limitations. First, the ADHD web portal collects few patient demographics, and therefore the representativeness of the sample is unknown. Additionally, patients are registered in the portal as a means of evaluation by a primary care provider; other than this, little is known about the context of ratings or what prompted the ADHD web portal registration. While this may create concerns related to generalizability of the results, the primary care provider's office is often the first place children present with concerns related to IA or HI (Zarin, Tanielian, Suarez, & Marcus, 1998; Zito et al., 1999). Similarly, aside from the Vanderbilt ADHD Rating Scales, no diagnostic information is available for enrolled patients, and the number of patients who were ultimately diagnosed with ADHD is unknown. There also may be concerns about whether these findings generalize to all ADHD referrals, since we found that patients who had teacher ratings had significantly higher parent ratings of IA symptoms than did patients who were excluded from the study because they were missing teacher ratings. In addition, the academic subject area of approximately 25% of the teachers included in the analyses was identified as *other*. Therefore, it is possible that the teachers who provided ratings for the patients in the study had varying degrees of structure in their classroom or observed the child in varied settings. Finally, it is possible that the match or mismatch between parent and teacher gender could impact their agreement (or lack thereof) on child behavior ratings; however, this question cannot be assessed, given the lack of demographic information about respondents. These limitations may impact the generalizability of these findings.

Future studies are needed to examine whether patient- and/or rater-related factors may moderate parent–teacher agreement. Variables such as time spent with the child, structure of the setting in which the child is observed, and parent–teacher communication are all potential variables that have been alluded to in the discussion of our results that may impact inter-rater agreement. Specific studies to examine these factors may further elucidate the causes of discordant ratings across parents and teachers. Another area in need of additional study within the inter-rater agreement literature is exploring parent–teacher agreement on ratings of children's impairment in daily functioning. Given that functional impairments (e.g., poor academic performance, peer difficulties), not ADHD symptoms, are often what leads families to seek assessment and treatment (Chronis et al., 2001), it would be interesting and clinically relevant to examine parent–teacher agreement on functional impairment outcomes.

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*, 213–232. doi:10.1037/0033-2909.101.2.213

Amador-Campos, J. A., Forns-Santacana, M., Guardia-Olmos, J., & Pero-Cebollero, M. (2006). *DSM–IV* attention-deficit/hyperactivity disorder symptoms: Agreement between informants in prevalence and factor structure at different ages. *Journal of Psychopathology and Behavioral Assessment, 28*, 23–32. doi:10.1007/s10862-006-4538-x

American Academy of Child and Adolescent Psychiatry. (2007). Practice parameters for the assessment and treatment of children and adolescents with attention-deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 894–921. doi:10.1097/chi.0b013e318054e724

American Academy of Pediatrics. (2011). ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/

hyperactivity disorder in children and adolescents. *Pediatrics, 128,* 1007–1022. doi:10.1542/peds.2011-2654

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Antrop, I., Roeyers, H., Oosterlaan, J., & Van Oost, P. (2002). Agreement between parent and teacher ratings of disruptive behavior disorders in children with clinically diagnosed ADHD. *Journal of Psychopathology and Behavioral Assessment, 24,* 67–73. doi:10.1023/A:1014057325752

Bard, D. E., Wolraich, M. L., Neas, B., Doffing, M., & Beck, L. (2013). The psychometric properties of the Vanderbilt Attention-Deficit Hyperactivity Disorder Diagnostic Parent Rating Scale in a community population. *Journal of Developmental & Behavioral Pediatrics, 34,* 72–82. doi:10.1097/DBP.0b013e31827a3a22

Burns, G. L., de Moura, M. A., Walsh, J. A., Desmul, C., Silpakit, C., & Sommers-Flanagan, J. (2008). Invariance and convergent and discriminant validity between mothers' and fathers' ratings of oppositional defiant disorder toward adults, ADHD-HI, ADHD-IN, and academic competence factors within Brazilian, Thai, and American children. *Psychological Assessment, 20,* 121–130. doi:10.1037/1040-3590.20.2.121

Burns, G. L., Gomez, R., Walsh, J. A., & de Moura, M. A. (2003). Understanding source effects in ADHD rating scales: Reply to DuPaul (2003). *Psychological Assessment, 15,* 118–119. doi:10.1037/1040-3590.15.1.118

Burns, G. L., Taylor, T. K., & Rusby, J. C. (2001a). *Child and Adolescent Disruptive Behavior Inventory 2.3: Parent Version*. Pullman, WA: Washington State University, Department of Psychology.

Burns, G. L., Taylor, T. K., & Rusby, J. C. (2001b). *Child and Adolescent Disruptive Behavior Inventory 2.3: Teacher Version*. Pullman, WA: Washington State University, Department of Psychology.

Burns, G. L., Walsh, J. A., Gomez, R., & Hafetz, N. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment, 18,* 452–457. doi:10.1037/1040-3590.18.4.452

Burns, G. L., Walsh, J. A., Servera, M., Lorenzo-Seva, U., Cardo, E., & Rodriguez-Fornells, A. (2013). Construct validity of ADHD/ODD rating scales: Recommendations for the evaluation of forthcoming *DSM–V* ADHD/ODD scales. *Journal of Abnormal Child Psychology, 41,* 15–26. doi:10.1007/s10802-012-9660-5

Chen, F. F., Sousa, K. H., & West, G. S. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12,* 471–492. doi:10.1207/s15328007sem1203_7

Choudhury, M. S., Pimentel, S. S., & Kendall, P. C. (2003). Childhood anxiety disorders: Parent-child (dis)agreement using a structured interview for the *DSM–IV*. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 957–964. doi:10.1097/01.CHI.0000046898.27264.A2

Chronis, M. A., Fabiano, G. A., Gnagy, E. M., Wymbs, B. T., Burrows-MacLean, L., & Pelham, W. E., Jr. (2001). Comprehensive, sustained behavioral and pharmacological treatment for attention-deficit/hyperactivity disorder: A case study. *Cognitive and Behavioral Practice, 8,* 346–359. doi:10.1016/S1077-7229(01)80008-0

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998). The Revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 257–268. doi:10.1023/A:1022602400621

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509. doi:10.1037/0033-2909.131.4.483

DuPaul, G. J. (2003). Assessment of ADHD symptoms: Comment on Gomez et al. (2003). *Psychological Assessment, 15,* 115–117. doi:10.1037/1040-3590.15.1.115

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., 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,* 436–444. doi:10.1037/1040-3590.9.4.436

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *The ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation*. New York, NY: Guilford Press.

Engel, N. A., Rodrigue, J. R., & Geffken, G. R. (1994). Parent-child agreement on ratings of anxiety in children. *Psychological Reports, 75,* 1251–1260. doi:10.2466/pr0.1994.75.3.1251

Epstein, J. N., Langberg, J. M., Lichtenstein, P. K., Kolb, R., & Simon, J. O. (2013). The myADHDportal.com Improvement Program: An innovative quality improvement intervention for improving the quality of ADHD care among community-based pediatricians. *Clinical Practice in Pediatric Psychology, 1,* 55–67. doi:10.1037/cpp0000004

Evans, S. W., Allen, J., Moore, S., & Strauss, V. (2005). Measuring symptoms and functioning of youth with ADHD in middle schools. *Journal of Abnormal Child Psychology, 33,* 695–706. doi:10.1007/s10802-005-7648-0

Froehlich, T. E., Lanphear, B. P., Epstein, J. N., Barbaresi, W. J., Katusic, S. K., & Kahn, R. S. (2007). Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of U.S. children. *Archives of Pediatrics and Adolescent Medicine, 161,* 857–864. doi:10.1001/archpedi.161.9.857

Gomez, R., Burns, G. L., Walsh, J. A., & de Moura, M. A. (2003). Multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychological Assessment, 15,* 3–16. doi:10.1037/1040-3590.15.1.3

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry, 7,* 125–130.

Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the Anxiety Disorders Interview Schedule for parents and children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 30–40. doi:10.1097/00004583-200301000-00008

Harpin, V. A. (2005). The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. *Archives of Disease in Childhood, 90*(Suppl. 1), i2–i7. doi:10.1136/adc.2004.059006

Hart, E. L., Lahey, B. B., Loeber, R., Applegate, B., & Frick, P. J. (1995). Developmental change in attention-deficit hyperactivity disorder in boys: A four-year longitudinal study. *Journal of Abnormal Child Psychology, 23,* 729–749. doi:10.1007/BF01447474

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Jensen, P. S., Xenakis, S. N., Davis, H., & Degroot, J. (1988). Child psychopathology rating scales and interrater agreement: II. Child and family characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 27,* 451–461. doi:10.1097/00004583-198807000-00013

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of Child Psychology and Psychiatry, 34,* 991–1006. doi:10.1111/j.1469-7610.1993.tb01103.x

Larsson, J. O., Larsson, H., & Lichtenstein, P. (2004). Genetic and environmental contributions to stability and change of ADHD symptoms between 8 and 13 years of age: A longitudinal twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 1267–1275. doi:10.1097/01.chi.0000135622.05219.bf

Little, D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341. doi:10.1207/s15328007sem1103_2

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis Group.

Mitsis, E. M., McKay, K. E., Schulz, K. P., Newcorn, J. H., & Halperin, J. M. (2000). Parent-teacher concordance for *DSM–IV* attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 308–313. doi:10.1097/00004583-200003000-00012

Molina, B. S. G., Smith, B. H., & Pelham, W. E. (2001). Secondary school teacher ratings of ADHD and ODD. *Journal of Abnormal Child Psychology, 29,* 71–82. doi:10.1023/A:1005203629968

Murray, D. W., Kollins, S. H., Hardy, K. K., Abikoff, H. B., Swanson, J. M., Cunningham, C., . . . Chuang, S. Z. (2007). Parent versus teacher ratings of attention-deficit/hyperactivity disorder symptoms in the Preschoolers With Attention-Deficit/Hyperactivity Disorder Treatment Study (PATS). *Journal of Child and Adolescent Psychopharmacology, 17,* 605–619. doi:10.1089/cap.2007.0060

Muthén, L. K., & Muthén, B. O. (1998–2012a). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998–2012b). *Mplus Version 7.1 language addendum*. Los Angeles, CA: Muthén & Muthén.

Papageorgiou, V., Kalyva, E., Dafoulis, V., & Vostanis, P. (2008). Differences in parents' and teachers' ratings of ADHD symptoms and other mental health problems. *European Journal of Psychiatry, 22,* 200–210. doi:10.4321/S0213-61632008000400003

Sayal, K., & Goodman, R. (2009). Do parental reports of child hyperkinetic disorder symptoms at school predict teacher ratings? *European Child & Adolescent Psychiatry, 18,* 336–344. doi:10.1007/s00787-009-0735-y

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 28–38. doi:10.1097/00004583-200001000-00014

Sibley, M. H., Pelham, W. E., Jr., Molina, B. S., Gnagy, E. M., Waschbusch, D. A., Garefino, A. C., . . . Karch, K. M. (2012). Diagnosing ADHD in adolescence. *Journal of Consulting and Clinical Psychology, 80,* 139–150. doi:10.1037/a0026577

Sollie, H., Larsson, B., & Morch, W. T. (2013). Comparison of mother, father, and teacher reports of ADHD core symptoms in a sample of child psychiatric outpatients. *Journal of Attention Disorders, 17,* 699–710. doi:10.1177/1087054711436010

Verhulst, F. C., Althaus, M., & Berden, G. F. (1987). The Child Assessment Schedule: Parent-child agreement and validity measures. *Journal of Child Psychology and Psychiatry, 28,* 455–466. doi:10.1111/j.1469-7610.1987.tb01766.x

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., . . . Lahey, B. B. (2012). Validity of *DSM–IV* attention deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of Abnormal Psychology, 121,* 991–1010. doi:10.1037/a0027347

Wolraich, M. L., Bard, D. E., Neas, B., Doffing, M., & Beck, L. (2013). The psychometric properties of the Vanderbilt Attention-Deficit Hyperactivity Disorder Diagnostic Teacher Rating Scale in a community population. *Journal of Developmental and Behavioral Pediatrics, 34,* 83–93. doi:10.1097/DBP.0b013e31827d55c3

Wolraich, M. L., Feurer, I. D., Hannah, J. N., Baumgaertel, A., & Pinnock, T. Y. (1998). Obtaining systematic teacher reports of disruptive behavior disorders utilizing *DSM–IV*. *Journal of Abnormal Child Psychology, 26,* 141–152. doi:10.1023/A:1022673906401

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. *Journal of Developmental and Behavioral Pediatrics, 25,* 41–47. doi:10.1097/00004703-200402000-00007

Wolraich, M. L., Lambert, W., Doffing, M. A., Bickman, L., Simmons, T., & Worley, K. (2003). Psychometric properties of the Vanderbilt ADHD Diagnostic Parent Rating Scale in a referred population. *Journal of Pediatric Psychology, 28,* 559–568. doi:10.1093/jpepsy/jsg046

Zarin, D. A., Tanielian, T. L., Suarez, A. P., & Marcus, S. C. (1998). Treatment of attention-deficit hyperactivity disorder by different physician specialties. *Psychiatric Services, 49*(2), 171.

Zito, J. M., Safer, D. J., dosReis, S., Magder, L. S., Gardner, J. F., & Zarin, D. A. (1999). Psychotherapeutic medication patterns for youths with attention-deficit/hyperactivity disorder. *Archives of Pediatrics and Adolescent Medicine, 153,* 1257–1263. doi:10.1001/archpedi.153.12.1257

Received August 28, 2012
Revision received July 22, 2014
Accepted July 31, 2014 ■