Parent and Teacher Ratings of Attention-Deficit/Hyperactivity Disorder Symptoms: Factor Structure and Normative Data

By: George H. DuPaul, Arthur D. Anastopoulos, Marley W. Watkins, Robert Reid, Matthew C. Lambert, Thomas J. Power

DuPaul, G. J., Reid, R., Anastopoulos, A. D., Lambert, M. C., Watkins, M. W., & Power, T. J. (2016). Parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. Psychological Assessment, 28(2), 214–225. https://doi.org/10.1037/pas0000166

Abstract:

Comprehensive assessment of attention-deficit/hyperactivity disorder (ADHD) symptoms includes parent and teacher questionnaires. The ADHD Rating Scale–5 was developed to incorporate changes for the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM–5; American Psychiatric Association, 2013). This study examined the fit of a correlated, 2-factor structure of ADHD (i.e., DSM–5 conceptual model) and alternative models; determined whether ADHD symptom ratings varied across teacher and child demographic characteristics; and presented normative data. Two samples were included: (a) 2,079 parents and guardians (1,131 female, 948 male) completed ADHD symptom ratings for children (N 2,079; 1,037 males, 1,042 females) between 5 and 17 years old (M 10.68; SD 3.75) and (b) 1,070 teachers (766 female, 304 male) completed ADHD symptom ratings for students (N 2,140; 1,070 males, 1,070 females) between 5 and 17 years old (M 11.53; SD 3.54) who attended kindergarten through 12th grade. The 2-factor structure was confirmed for both parent and teacher ratings and was invariant across child gender, age, informant, informant gender, and language. In general, boys were higher in symptom frequency than girls; older children were rated lower than younger children, especially for hyperactivity–impulsivity; and non-Hispanic children were rated higher than Hispanic children. Teachers also rated non-Hispanic African American children higher than non-Hispanic White, Asian, and Hispanic children. Non-Hispanic White teachers provided lower hyperactivity–impulsivity ratings than non-Hispanic, African American, and Hispanic teachers. Normative data are reported separately for parent and teacher ratings by child gender and age. The merits of using the ADHD Rating Scale–5 in a multimodal assessment protocol are discussed.

Keywords: attention-deficit/hyperactivity disorder | ADHD | assessment | parent ratings | teacher ratings

Article:

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that involves exhibition of developmentally inappropriate levels of inattention (IA) and/or hyperactivity– impulsivity (HI; American Psychiatric Association [APA], 2013). Epidemiological studies consistently indicate that approximately 5% of children and adolescents worldwide have ADHD (Polanczyk, de Lima, Horta, Biederman, & Rohde, 2007), with an even greater percentage (11%) of children in the United States having received the diagnosis (based on parent report) from community practitioners (Visser et al., 2014). ADHD symptoms begin during childhood; continue across the life span for most individuals; and are associated with significant impairment in academic, social, emotional, and/or occupational functioning (APA, 2013; Barkley, 2015). Thus, measures that can screen for and assess symptoms and related impairment are critical for successful treatment of this disorder.

A comprehensive approach to the diagnostic assessment of ADHD in children and adolescents includes diagnostic interview(s) with parent(s) and teacher(s), behavior rating scales, direct observation of school behavior (when feasible), and collection of data that will establish symptom-related impairment in academic and social functioning (Anastopoulos & Shelton, 2001; Barkley, 2015; DuPaul & Stoner, 2014; Pelham, Fabiano, & Massetti, 2005). An integral component of comprehensive assessment is the use of behavior rating scales completed by parents and teachers who observe child and adolescent behavior in home and school settings. Parent and teacher ratings are critically important because ADHD symptoms may not be evident in novel and controlled situations such as clinic-based evaluations (Power, 1992). Several behavior rating scales have been developed to obtain parent and teacher reports of ADHD symptom frequency and/or severity, including the ADHD Rating Scale–IV (DuPaul, Power, Anastopoulos, & Reid, 1998), the ADHD Symptoms Rating Scale (Holland, Gimpel, & Merrell, 1998), the Conners Rating Scale (Conners–3; Conners, 2008), the SNAP–IV Rating Scale (Swanson, Nolan, & Pelham, 1992), and the Vanderbilt ADHD Rating Scale (Wolraich, Hannah, Baumgaertel, & Feurer, 1998). These narrowband scales have demonstrated adequate to strong reliability and validity for screening, diagnosis, and treatment evaluation (Pelham et al., 2005).

Currently available ADHD rating scales, which were developed in accordance with guidelines set forth in the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; DSM– IV–TR; APA, 2000), are limited for at least two reasons. First, in contrast with the DSM–IV–TR, the new DSM–5 (APA, 2013) guidelines for ADHD now include additional wording for ADHD symptoms for use when evaluating adolescents. Thus, to be developmentally appropriate, rating scale items should include symptom descriptions that are appropriate for both children and adolescents. Second, previously published ADHD rating scales have normative data based on DSM–IV–TR criteria that were collected more than 10 years ago and may be outdated for use with current population cohorts. Thus, to address these limitations, the ADHD Rating Scale–5 was developed to incorporate symptom wording changes consistent with the DSM–5 and to obtain current normative data for symptom ratings based on DSM–5 criteria for a nationally representative sample of children and adolescents.

It is also important to consider the structure of ADHD symptoms, as this has direct implications for our conceptual understanding of the disorder, as well as for the development of questionnaire subscales. The DSM–5 criteria for ADHD include separate listings of IA and hyperactive–impulsive symptoms based primarily on factor analytic findings for DSM–IV criteria that supported a correlated two-factor structure (e.g., Baumgaertel, Wolraich, & Dietrich, 1995; DuPaul et al., 1997; Molina, Smith, & Pelham, 2001; Pillow, Pelham, Hoza, Molina, & Stultz, 1998). However, results from prior investigations are limited in at least two ways. First, the factor structure of DSM–5 ADHD symptoms is unknown, especially for a general population (i.e., nonclinical sample). All prior studies have been conducted using DSM–IV criteria, and most have collected data for clinical samples. Second, no prior study has comprehensively examined the structural invariance of the DSM–5 model across respondents (parents and teachers), age groups (children and adolescents), and child gender. It is possible that symptom structure could vary as a function of who is reporting on the symptoms, as well as the age and gender of the individual being rated.

Consequently, the purpose of this study was fourfold: (a) to examine the fit of the correlated, two-factor structure of ADHD (i.e., the DSM–5 conceptual model) and alternative models; (b) to assess whether factor structure varies as a function of respondent (parent, teacher), age and gender of individuals being assessed, or language used (English and Spanish parent ratings); (c) within this DSM–5 factor structure, to determine whether ADHD symptom ratings vary as a function of teacher (e.g., race and ethnicity) and child demographic characteristics (e.g., age, sex, race, ethnicity); and (d) to present normative data for the ADHD Rating Scale–5 consistent with the DSM–5 symptom structure. It was hypothesized that the two-factor DSM–5 model would provide adequate fit and be superior to alternative models (i.e., one factor, three factor, bifactor). Further, it was hypothesized that the two-factor structure would be invariant across respondent, child age, child gender, and Spanish–English versions. Based on prior findings for DSM– IV–TR measures, ADHD symptom ratings were expected to be higher for males, younger children, and individuals from racial and ethnic minority backgrounds.

Method

Participants

Two separate samples were included in this study. One sample included 2,079 parents and guardians (1,131 female, 948 male) who completed ADHD symptom ratings for one of their children selected at random (see the Procedures section). Parents and guardians were predominantly White non-Hispanic (64.1%) and ranged in age from 20 to 77 years old (M =

41.57; SD = 8.23). Most parents were married (79.7%), had at least a high school education or greater (89.9%), and were employed (72.3%). Household income ranged from less than $5,000 (2.7%) to $175,000 or more (5.7%), with median income between $60,000 and $74,999 and mean income between $50,000 and $59,999. The parent sample was recruited from all regions of the United States and included households from both metropolitan (86.4%) and nonmetropolitan (13.6%) locations. English was spoken in most (89.4%) households. The children (N = 2,079; 1,037 males, 1,042 females) rated by the parents ranged in age from 5 to 17 years old (M = 10.68; SD = 3.75). Children were from White non-Hispanic (53.9%), Black non-Hispanic (13.1%), Asian non-Hispanic (5.7%), Hispanic (23.4%), and other (3.9%) backgrounds.

The second sample included 1,070 teachers (766 female, 304 male) who completed ADHD symptom ratings; each teacher rated two randomly selected students (one male, one female) on their class rosters (see the Procedures section). Teachers were predominantly White non-Hispanic (87.3%) and reported a mean of 17.88 years of teaching experience (SD = 10.7). The teacher sample was recruited from all regions of the United States and included general (83.3%) and special education (16.4%) teachers. The students (N = 2,140; 1,070 males, 1,070 females) rated by the teachers ranged in age from 5 to 17 years old (M = 11.53; SD = 3.54) and attended kindergarten through 12th grade. Most students attended general education classrooms (83.2%) and were from White nonHispanic (54.8%), Black non-Hispanic (12.7%), other non Hispanic (7.0%), Hispanic (24%), or biracial non-Hispanic (1.5%) backgrounds.

Measures

Demographic characteristics. Parents and teachers reported their gender, age, race, and ethnicity. Parents also reported marital status, level of education, employment status, household income, and language spoken in the household. Teachers also reported years of teaching experience and primary teaching assignment (i.e., general education or special education). Parents and teachers provided information about the children they rated, including gender, age, race, and ethnicity. In addition, teachers reported the grade and primary classroom placement (i.e., general education or special education) for these students.

ADHD symptom ratings. Parents and teachers reported the frequency with which each child displayed the 18 symptomatic behaviors of ADHD (see Figure 1) using the ADHD Rating Scale–5 Home and School versions, respectively. With the permission of the APA, items were created based on the wording of ADHD symptoms from the DSM–5. Parents and teachers indicated the frequency of each behavior on a 4-point Likert scale, including 0 (never or rarely), 1 (sometimes), 2 (often), and 3 (very often). For the Home Version, parents were asked to select the number that best described their child’s behavior over the previous 6 months. For the School Version, teachers were also asked to select the number that best described the student’s behavior over the past 6 months or since the beginning of the school year. Otherwise, the Home and School versions were identical in item wording and format. For students in Grades 6 through 12 (i.e., adolescents age 11 and older), additional wording (from the DSM–5) was provided for some items to make these developmentally relevant. For example, the IA item “has difficulty sustaining attention in tasks or play activities” was amended to include the following parenthetical text: “(e.g., has difficulty remaining focused during lectures, conversations, or lengthy reading).” The nine IA items were listed separately from the nine HI items. For the hypothesized two-factor model (based on DSM–5 criteria), IA and HI factor scores were derived separately by summing the nine item scores for each factor. Alternatively, the single-factor model involved summing ratings for all 18 items into one total score, while the three-factor model involved summing nine items for IA, six items for hyperactivity, and three items for impulsivity.

Parents whose primary language was Spanish (n 236; 11.4%) completed a version of the ADHD Rating Scale–5 that included 18 symptom items using wording from the Spanish edition of the DSM–5 (APA, 2013). The translation process involved (a) initial translation into Spanish; (b) independent review by two specialists trained in language elements of diverse cultures; (c) collaboration between independent reviewers; and (d) involvement of a senior translator/researcher, if necessary, to resolve differences.

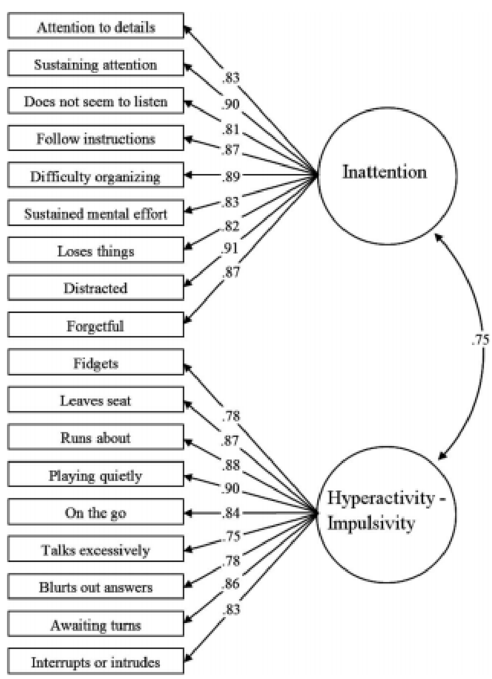

Figure 1. Two-factor model: parent report on English version of ADHD Rating Scale-5.

Procedure

Parents were recruited through GfK, a national research firm (see Table 1). All parent respondents (N 2,079) were recruited through the GfK KnowledgePanel to provide a sample of children and adolescents that was representative of the U.S. population in terms of race, ethnicity, geographic region, and family income (see Table 2). KnowledgePanel is a large, national, probability-based panel that provides online research for measurement of public opinion, attitudes, and behavior. Panelists were selected using address-based sampling (ABS) that allows probability-based sampling of addresses from the U.S. Postal Service's Delivery Sequence File. Individuals residing at randomly sampled addresses were invited to join KnowledgePanel through a series of mailings (in English and Spanish); nonresponders were phoned when a telephone number could be matched to the sampled address. Household members who were randomly selected indicated their willingness to join the panel by returning a completed acceptance form in a postage-paid envelope, calling a toll-free hotline and speaking to a bilingual recruitment agent, or accessing a dedicated recruitment website. If more than one child between the ages of 5 and 17 was present in a given household, then parents were asked to provide ratings for one randomly selected child such that the number of cases was balanced across gender and age range.

Teacher data were collected via two national research firms: GfK KnowledgePanel and e-Rewards. Initially, 1,509 teachers on the KnowledgePanel were assigned to complete ratings (see Table 1). To obtain the desired sample size of 2,000 students, additional teachers were recruited through e-Rewards Market Research; e-Reward panelists are selected based on having a relationship with a business (e.g., Pizza Hut, Hertz, Macy's). A double opt-in is required; panelists must reply to the initial e-mail invitation and then to a follow-up confirmation e-mail. Potential panelists' physical addresses are then verified against postal records. All respondents are required to have a valid and unique e-mail address. At present, e-Rewards has over 3,000,000 active panelists. Respondents answer a profiling questionnaire when enrolling and provide information regarding employment status. The e-Rewards respondents indicated employment as a full-time, regularly employed (i.e., not substitute) teacher. To ensure equal gender representation, all teachers were asked to provide symptom ratings for one randomly selected boy and one randomly selected girl on their class roster. Each student selected was based on a randomly generated number provided in the instructions. Thus, for example, the teacher might be asked to select the seventh girl on the class roster. This procedure was used for both students rated. Secondary school teachers were instructed to provide ratings for one randomly selected male and one randomly selected female in a randomly selected class. Further, the sample was recruited such that the number of cases was balanced across age and grade range and was representative of the U.S. child population in terms of race– ethnicity, geographic region, and age–sex (see Table 3).

Parent and teacher ratings were completed through a web-based survey that took approximately 5 min and 9 min (i.e., less than 5 min per student), respectively. Respondents received small stipends (less than $5) for completing ratings. If respondents left one or more items blank, they were prompted to complete missing items. Thus, complete data sets were produced for 99% of child ratings. All ratings were collected during April and May 2014.

Table 1. Recruitment Procedures for Teacher and Parent Ratings

Teachers		Parents
GfK Address-based national probability sampling of individuals employed as teachers invited to join knowledge panel (N = 1,509)	e-Reward e-Reward member indicating employment as teacher invited to participate (N = 12,610)	GfK Address-based national probability sampling of parents–caregivers invited to join knowledge panel (N = 4,219)
Completed ratings (N = 1,019)	Completed ratings (N = 1,399)	Completed ratings (N = 2,708)
Qualified based on demographic quotas (student grade, race–ethnicity, geographic region; N = 474)	Qualified based on demographic quotas (student grade, race–ethnicity, geographic distribution; N = 596)	Qualified based on demographic quotas (student grade, race–ethnicity, geographic region; N = 2,079)

Table 2. Parent Ratings: Sample Versus U.S. Census

Category	Percent sample	Percent census
Race–ethnicity		
White, non-Hispanic	53.9	53.55
Black, non-Hispanic	13.1	13.75
Hispanic	23.4	23.30
Other, non-Hispanic	5.7	5.85
Two or more races, non-Hispanic	3.9	3.55
Region		
Northeast	16.6	16.57
Midwest	21.7	21.50
South	37.5	37.61
West	24.4	24.32
Income level		
Under $25,000	18.3	18.63
$25,000–$49,000	21.7	21.71
$50,000–$74,999	24.9	17.65
$75,000 and over	35.1	42.01
Age–sex		
5- to 7-year-old boy	11.73	11.64
5- to 7-year-old girl	11.15	11.13
8- to 10-year-old boy	11.44	11.43
8- to 10-year-old girl	10.87	11.09
11- to 13-year-old boy	11.78	11.79
11- to 13-year-old girl	11.21	11.17
14- to 17-year-old boy	15.87	16.21
14- to 17-year-old girl	15.39	15.54

Note. U.S. population benchmarks. March 2013 Current Population Survey supplement data

Table 3. Teacher Ratings: Sample Versus U.S. Census

Category	Teacher sample	Percent census
Race–ethnicity		
White, non-Hispanic	54.8	53.55
Black, non-Hispanic	12.7	13.75
Hispanic	24.0	23.30
Other, non-Hispanic	7.0	5.85
Two or more races, non-Hispanic	1.5	3.55
Region		
Northeast	23.2	16.81
Midwest	27.1	21.68
South	26.6	37.53
West	23.1	23.97
Age–sex		
5- to 7-year-old boy	11.1	11.64
5- to 7-year-old girl	10.6	11.13
8- to 10-year-old boy	11.8	11.43
8- to 10-year-old girl	11.4	11.09
11- to 13-year-old boy	12.3	11.79
11- to 13-year-old girl	11.5	11.17
14- to 17-year-old boy	15.9	16.21
14- to 17-year-old girl	15.1	15.54

Note. U.S. population benchmarks. March 2013 Current Population Survey supplement data

Data Analyses

SPSS Version 21was used to compute descriptive statistics and univariate factorial analyses of covariance (ANCOVAs). Mplus Version 7.11 (Muthén & Muthén, 1998 –2014) was used to fit confirmatory factor analysis (CFA) models. The focus of the factor analysis was to examine the fit of the theoretical two-factor DSM–5 model. As a basis for comparison, three alternative factor models were also fit to the data: (a) a single-factor model; (b) a three-factor model (with separate Inattention, Hyperactivity, and Impulsivity factors) owing to prior conceptualizations of the structure of ADHD symptoms—that is, the Diagnostic and Statistical Manual of Mental Disorders (3rd ed.; DSM–III; APA, 1980); and (c) a bifactor model including a General ADHD factor, as well as IA and HI group factors, as found in prior studies (e.g., Ullebø, Breivik, Gillberg, Lundervold, & Posserud, 2012). Parent ratings on the English version of the ADHD Rating Scale–5 were used in testing all CFA models.

Bifactor versions of the two-factor model were generated to obtain an additional perspective on scale structure through a partition of the item response variance into common sources (Reise, 2012). Bifactor models posit one general factor and several group factors, all orthogonal to each other. Thus, the bifactor model allows each factor to directly influence the items independent of the other factors (DeMars, 2013) and can complement a correlated factors model "by evaluating whether item response variance is due to a general construct versus group factors" (Brouwer, Meijer, & Zevalkink, 2013, p. 138). Modeling of a General ADHD factor, as well as IA and HI group factors, is similar to what has been recommended by other researchers (e.g., Ullebø et al., 2012) and is congruent with theoretical models (Burns, de Moura, Beauchaine, & McBurnett, 2014).

Because items were measured on a 4-point Likert-type scale, we treated the ratings as ordinal rather than continuous indicators of the latent factors. Accordingly, we used weighted least squares with mean and variance adjustments (WLSMV; robust WLS) to estimate each model, and the factors were scaled using a fixed mean and variance approach. All CFA models were specified without correlated residual variances between items. Missing data for the CFA models were minimal (0.01%) and were excluded from the analysis by using a pairwise–present method as is default in Mplus when using the WLSMV estimator.

The indicators used to assess goodness of fit were the comparative fit index (CFI; Bentler, 1990), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA; Steiger & Lind, 1980) at its 90% confidence interval (CI). CFI and TLI are comparative fit indices representing the degree of improvement over the worst-fitting model (Boomsma, 2000). Both indices are scaled from 0 to 1, with values closer to 1 indicating better fit. An acceptable-fitting model has CFI and TLI values greater than or equal to 0.95 (Browne & Cudeck, 1993). RMSEA represents the degree of model misfit and is reported on a scale of 0 to 1; values closer to zero indicate better fit, with values close to .06 balancing Type I and Type II error rates (Hu & Bentler, 1999). In addition to examining the point estimate, the 90% CI was also used to evaluate misfit, with upper limits lower than .08 representing acceptable fit. The chi-square difference test (2) and CFI differences (CFI) were computed to evaluate the fit of nested models (e.g., the one-factor vs. the two-factor model). Nonsignificant tests or differences in CFI less than .01 indicate that the fit of the two models being compared is statistically equivalent.

Measurement invariance was tested for the two-factor model using the multiple-group CFA framework for detecting significant differences in measurement parameters (i.e., factor loadings and thresholds) by child gender, ethnicity, and age (for age groups 5–7, 8 –10, 11–13, and 14 –17 years); between the rater’s gender; and between the Spanish and English versions of the parent scale. The age groups were selected to represent children’s ages at important school levels (i.e., early elementary school, late elementary school, middle school, and high school). Before formally testing invariance, the fit to the two-factor model was evaluated for each group separately. Then, measurement invariance was assessed by comparing the fit of a multiple-group configural invariance model (i.e., factor loadings and thresholds allowed to vary across groups with fixed factor means and scaling parameters for the latent response variables; see Muthén & Muthén, 1998 –2014, p. 485) with the fit of a strong invariance model (i.e., factor loadings and thresholds constrained across groups with the factor means and scaling parameters fixed in one group and estimated in the other). When using WLSMV to estimate a model with categorical factor indicators, testing the intermediate step of metric invariance (i.e., “weak measurement invariance”) is not commonly recommended because factor loadings and thresholds are most interpretable when either both fixed or both estimated (Millsap, 2011). The Satorra– Bentler chi-square difference test and differences in CFI values (CFI .01) were used to indicate whether the constraints were tenable (i.e., the measurement parameters were statistically and substantively equivalent; Cheung & Rensvold, 2002).

Results

Factor Analysis

Table 4 reports the goodness-of-fit indicators for the parent and teacher CFA models. The single-factor model did not exhibit acceptable fit. For both parent and teacher ratings, the two-factor model demonstrated acceptable fit (CFI ≥.96, TLI ≥.95, RMSEA close to .06), with the upper limit of the RMSEA CI within the acceptable range. The two latent factors were highly correlated, .75, 95% CI [.73, .77] and .80, CI [.78, .82], for the parent and teacher models, respectively. The three-factor model demonstrated a statistically significant improvement in fit over the two-factor model for the parent, $\Delta\chi 2(2)$ = 135.55, p $\leqslant$ .001, and teacher ratings, $\Delta\chi 2(2)$ = 164.56, p $\leqslant$ .001; however, the improvement in fit is of limited substantive importance, as indicated by the small ΔCFI ($\leqslant$.01) between the models. Graphical models complete with fully standardized factor loadings and the correlation between the latent factors for the favored two-factor model are presented in Figures 1 and 2.

As expected, the bifactor versions of parent- and teacher- correlated factor models were good fits to the data (see Table 4). A strong General ADHD factor, a distinct IA factor, and a weak HI factor emerged for both parents and teachers (e.g., Burns et al., 2014; Normand, Flora, Toplak, & Tannock, 2012; Toplak et al., 2012; Ullebø et al., 2012). In particular, all hyperactivity items strongly loaded on the General ADHD factor and were only weakly related to the HI factor for both parent and teacher ratings. For parents, 74% of the explained common variance was due to the General ADHD factor, 21% was attributed to the IA factor, and 5% was attributed to the HI factor. Omega hierarchical (McDonald, 1999), a model-based reliability measure that does not rely on the assumptions of the coefficient alpha, was .85 for the General ADHD factor, .41 for the IA factor, and .03 for the HI factor. For teacher ratings, 79% of the explained common variance was due to the General ADHD factor, 17% was attributed to the IA factor, and 4% was attributed to the HI factor. Omega hierarchical was .89 for the General ADHD factor, .33 for the IA factor, and .03 for the HI factor.

Given concerns about the weak HI group factor emerging from the bifactor analyses, as well as the conceptual and clinical value of the DSM–5 two-factor model, subsequent analyses (i.e., examination of effects of child and teacher demographic characteristics, development of normative data) were based upon the two-factor solution: Scores were derived for the IA factor (nine items) and HI (nine items) factor. In addition, a total score (all 18 items) was derived given its utility for clinical assessment, especially for evaluating ADHD, combined presentation.

Table 4. Confirmatory Factor Analysis Results for Parent and Teacher Ratings of Attention-Deficit/Hyperactivity Disorder (ADHD) Symptoms

	df	χ^2	CFI	TLI	RMSEA [95% CI]	$\Delta\chi^2$ (df)
Parents (N = 1,840)[a]						
Single-factor model	135	3,817.96	0.929	0.919	0.122 [.118, .125]	—
Two-factor model	134	1,262.77	0.978	0.975	0.068 [.064, .071]	355.75 (1)*
Bifactor version	117	464.07	0.993	0.991	0.040 [.036, .044]	
Three-factor model	132	901.86	0.985	0.983	0.056 [.053, .060]	135.55 (2)*
Teachers (N = 2,140)						
Single-factor model	135	4,998.69	0.966	0.962	0.130 [.127, .133]	—
Two-factor model	134	1,704.69	0.989	0.988	0.074 [.071, .077]	438.30 (1)*
Bifactor version	117	936.05	0.994	0.993	0.057 [.054, .061]	
Three-factor model	132	1,258.05	0.992	0.991	0.063 [.060, .066]	164.56 (2)*

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA= root mean square error of approximation; CI confidence interval.

[a] Data from English version of ADHD Rating Scale–5 only.

*$p \leqslant .001$.

Measurement invariance by rater gender. The two-factor model was tested separately for male and female raters, and the model fit the data acceptably for both groups (see Table 5). Multiple-group models were then fit to test measurement invariance across rater gender (see Table 6). For parent ratings, there were no appreciable differences in model fit between the configural invariance model and the strong invariance model, as indicated by the small change in CFI values (CFI .001) and a nonsignificant $\Delta\chi 2$ test, $\Delta\chi 2_{(50)}$ = 64.90, $p \succ .05$. For teacher ratings, there were only small differences in the measurement properties between genders, as indicated by no change in CFI values (ΔCFI $\leqslant$ .001) and small, albeit significant, $\Delta\chi 2$ values, $\Delta\chi 2_{(50)}$ = 114.06, $p \leqslant .001$. Such small differences in model fit indicate that the measurement parameters are substantively equivalent across gender for both parent and teacher ratings.

Measurement invariance by child gender, ethnicity, and age. When the two-factor model was fit to each group separately, RMSEA estimates and/or the upper limit of the RMSEA CI were, on occasion, greater than .08, indicating a slightly strained fit to the data (see Table 5). Ratings of male students and African American students, as well as teacher ratings for the 5–7, 8–10, and 11–13 age groups, exhibited a slight misfit to the data (see Table 5); however, CFI and TLI values exhibited a close fit across gender, race– ethnicity, and age groups. In general, measurement parameters and goodness-of-fit indicators did not differ substantively between males and females, between race and ethnic groups, or between age groups (see Table 6). Although all of the chi-square difference tests were statistically significant and indicated minor differences in measurement parameters, differences in CFI values did not suggest meaningful differences; all CFI values were less than or equal to .002.

Measurement invariance by language. Ratings from the English and Spanish versions of the parent rating scale both fit the two-factor model acceptably (see Table 5), and measurement parameters (of parent ratings) were relatively invariant across the English and Spanish versions of the assessment (see Table 6). Although the chi-square difference test indicated a statistically significant variation between the versions, (50) 2 87.20, p .01, the differences were of a small magnitude, as indicated by the minor CFI value (.003). Thus, data from the English and Spanish versions of the assessment were combined for analyses of gender, age, and ethnic differences.

Table 5.Two-Factor Confirmatory Factor Analysis (CFA) Results and Internal Consistency Reliability (Alpha) Coefficients by Subgroup

		Two-factor CFA				Alpha coefficient [95% CI]	
	N	χ^2	CFI	TLI	RMSEA [95% CI]	IA	HI
Parents[a]							
Child age							
5–7	475	394.47	0.981	0.978	0.064 [.057, .071]	.93 [.93, .94]	.91 [.90, .92]
8–10	442	315.36	0.988	0.986	0.053 [.045, .061]	.93 [.93, .94]	.92 [.91, .93]
11–13	442	475.37	0.979	0.976	0.076 [.069, .083]	.95 [.94, .96]	.90 [.89, .91]
14–17	481	349.48	0.984	0.981	0.058 [.050, .065]	.94 [.94, .95]	.89 [.87, .90]
Child gender							
Male	920	926.93	0.974	0.970	0.080 [.075, .085]	.94 [.94, .95]	.92 [.91, .92]
Female	920	467.82	0.985	0.983	0.052 [.047, .057]	.93 [.93, .94]	.90 [.89, .91]
Child race–ethnicity							
White	1,225	827.44	0.981	0.979	0.065 [.061, .069]	.94 [.94, .95]	.91 [.90, .92]
African American	142	254.36	0.978	0.975	0.080 [.065, .094]	.93 [.92, .95]	.92 [.90, .93]
Hispanic	254	283.94	0.977	0.973	0.066 [.056, .077]	.94 [.93, .95]	.90 [.88, .92]
Rater gender							
Male	868	565.46	0.980	0.977	0.061 [.056, .066]	.94 [.93, .94]	.91 [.90, .91]
Female	972	822.59	0.978	0.975	0.073 [.068, .078]	.94 [.94, .95]	.91 [.91, .92]
Assessment language							
English	1,840	1,262.77	0.978	0.975	0.068 [.064, .071]	.94 [.94, .94]	.91 [.90, .92]
Spanish	236	202.69	0.985	0.983	0.047 [.033, .059]	.94 [.92, .95]	.88 [.86, .91]
Teachers							
Child age							
5–7	348	439.90	0.991	0.990	0.081 [.073, .089]	.96 [.96, .97]	.95 [.94, .96]
8–10	559	589.39	0.988	0.986	0.078 [.072, .084]	.96 [.96, .97]	.94 [.93, .95]
11–13	510	556.60	0.988	0.987	0.079 [.072, .085]	.96 [.96, .97]	.94 [.93, .94]
14–17	723	475.46	0.991	0.990	0.059 [.054, .065]	.96 [.95, .96]	.94 [.94, .95]
Child gender							
Male	1,070	1,024.57	0.988	0.986	0.079 [.074, .083]	.96 [.96, .96]	.95 [.94, .95]
Female	1,070	761.93	0.989	0.988	0.066 [.062, .071]	.96 [.95, .96]	.93 [.92, .94]
Child race–ethnicity							
White	1,173	901.63	0.991	0.989	0.070 [.066, .074]	.96 [.96, .97]	.95 [.94, .95]
African American	271	384.40	0.986	0.984	0.083 [.073, .093]	.96 [.95, .97]	.94 [.93, .95]
Hispanic	514	522.66	0.988	0.987	0.075 [.068, .082]	.96 [.96, .97]	.95 [.94, .95]
Rater gender							
Male	608	467.43	0.993	0.992	0.064 [.058, .070]	.96 [.96, .97]	.95 [.94, .96]
Female	1,532	1,337.57	0.988	0.986	0.077 [.073, .080]	.96 [.96, .96]	.94 [.94, .95]

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; IA = inattention; HI = hyperactivity–impulsivity. Each model has 134 degrees of freedom.

[a] Data from English version of ADHD Rating Scale–5 only.

Gender, Age, and Ethnic Differences

Sum scores for parent and teacher scales were computed with unit weights (Wainer, 1976). A 2 (Child Gender) 4 (Child Age Groups 5–7, 8 –10, 11–13, and 14 –17) 3 (Child Race–Ethnicity: non-Hispanic White, non-Hispanic African American, and Hispanic) factorial ANCOVA was used to assess differences in sum scores for parent ratings on each factor across demographic groups while controlling for household income. A 2 4 3 analysis of variance (ANOVA) was used to assess differences for teacher ratings; measures of household income were not available as a covariate for teacher ratings. In addition, a 3 (Student Race– Ethnicity: non-Hispanic White, non-Hispanic African American, and Hispanic) x 3 (Teacher Race–Ethnicity: non-Hispanic White, non-Hispanic African American, and Hispanic) ANCOVA was used to estimate the interaction between the two factors while controlling for student gender and age. For these sets of analyses, the alpha level was set at .01 to account for the number of tests conducted. Cohen's d was computed for specific comparisons between two groups (i.e., gender) and partial eta squared (P 2) for comparisons between three or more groups (i.e., ethnicity and age). Cohen's d estimates for the parent ratings were computed based on covariateadjusted means and the mean square error of the ANCOVA. Partial eta squared can be interpreted as the proportion of variance explained by the factor after controlling for the other factors in a factorial ANCOVA or ANOVA.

Parent ratings. There were significant two-way (Gender x Age) interaction effects for the total score, $F(3, 1{,}882) = 5.25$, $p \leqslant .01$, $\eta 2P = .008$, and the HI score, $F(3, 1{,}882) = 5.84$, $p \leqslant .01$, $\eta 2P = .009$. For the Gender x Age interactions, post hoc tests indicated that males received higher scores compared to females for the youngest three age groups; however, females in the 14 –17 age group received slightly higher scores than males for both scores. The main effects of gender were significant for the total score, $F(1, 1{,}882) = 18.53$, $p \leqslant .001$, d = 0.17, 95% CI [0.08, 0.25], IA score, $F(1, 1{,}883) = 13.75$, $p \leqslant .001$, d = 0.21, CI [0.12, 0.30], and HI score, $F(1, 1{,}882) = 18.20$, $p \leqslant .001$, d = 0.23, CI [0.14, 0.32], where males received higher scores than females. There were significant main effects of age on the total score, $F(3, 1{,}882) = 4.04$, $p \leqslant .01$, $\eta 2P = .006$, and the HI score, $F(3, 1{,}882) = 18.38$, $p \leqslant .001$, $\eta 2P = .028$. Post hoc tests revealed that the youngest children (aged 5–7) received higher scores than children in the oldest age group (aged 15–17) on the total score and that children in the oldest group received significantly lower scores than children in the two youngest groups (aged 5–10) on HI. Lastly, the main effects of race–ethnicity were significant for the total score, $F(2, 1{,}882) = 7.59$, $p \leqslant .01$, $\eta 2P = .008$, IA score, $F(2, 1{,}883) = 6.12$, $p \leqslant .01$, $\eta 2P = .006$, and HI score, $F(2, 1{,}882) = 7.34$, $p \leqslant .001$, $\eta 2P = .008$. For the total score, non-Hispanic African American, d = 0.22, 95% CI [0.07, 0.37], and nonHispanic White children, d = 0.22, CI [0.11, 0.33], had higher scores than Hispanic children; for the IA score, non-Hispanic White children received higher scores than Hispanic children, d = 0.21, CI [0.10, 0.31]; and for the HI score, non-Hispanic African American, d = 0.25, CI [0.11, 0.40], and non-Hispanic White children, d = 0.20, CI [0.09, 0.31], received higher scores than Hispanic children.

Teacher ratings. There were no significant interaction effects for teacher ratings. The main effects of gender were significant for the total score, $F(1, 1{,}934) = 73.00$, $p \leqslant .001$, d = 0.47, 95% CI [0.38, 0.56], IA score, $F(1, 1{,}934) = 57.63$, $p \leqslant .001$, d = 0.42, CI [0.33, 0.51], and HI score, $F(1, 1{,}934) = 67.35$, $p \leqslant .001$, d = 0.45, CI [0.36, 0.54], where males received higher scores than females. There were significant main effects of age on the total score, $F(3, 1{,}934) = 14.69$, $p \leqslant .001$, $\eta 2P = .022$, IA score, $F(3, 1{,}934) = 10.97$, $p \leqslant .001$, $\eta 2P = .017$, and HI score, $F(3, 1{,}934) = 21.35$, $p \leqslant .001$, $\eta 2P = .032$. Post hoc tests revealed that the youngest children received higher scores than children in the oldest two age groups on the total score; the youngest three age groups received higher scores than the oldest age group on the IA subscale; and children in the oldest age group received significantly lower scores than all of the other children on HI.

Lastly, the main effects of race– ethnicity were significant for the total score, $F(2, 1{,}934) = 21.51$, $p \leqslant .001$, $\eta 2P = .022$, IA score, $F(2, 1{,}934) = 20.51$, $p \leqslant .001$, $\eta 2P = .021$, and HI score, $F(2, 1{,}934) = 16.35$, $p \leqslant .001$, $\eta 2P = .017$. For all three scores, non-Hispanic African American children received higher scores than non-Hispanic White children (total: d = 0.40, 95% CI [0.27, 0.53]; IA: d = 0.37, CI [0.24, 0.50]; HI: d = 0.38, CI [0.25, 0.51]) and Hispanic children (total: d = 0.28, CI [0.13, 0.42]; IA: d = 0.22, CI [0.08, 0.37]; HI: d = 0.29, CI [0.14, 0.44]). Scores for non-Hispanic White and Hispanic children did not differ significantly for any of the teacher ratings.

For the 3 x 3 ANCOVA, there was a significant two-way interaction between student race– ethnicity and teacher race– ethnicity for the HI score, $F(4, 1{,}740) = 3.77$, $p \leqslant .01$, $\eta 2P = .005$ (see Figure 3). Post hoc tests revealed that there were no differences between non-Hispanic White and Hispanic teachers' ratings of non-Hispanic White and Hispanic students; however, African American teachers rated Hispanic students significantly higher than non-Hispanic White students (d = 0.84, 95% CI [0.38, 1.30]). The interaction for the total score did not meet the adjusted significance level, $F(4, 1{,}740) = 2.83$, $p = .049$, $\eta 2P = .005$, but teachers exhibited the same pattern of ratings as for the HI score. There was a significant main effect of teacher race– ethnicity for the HI score, $F(2, 1{,}740) = 10.64$, $p \leqslant .01$, $\eta 2P = .012$, with non-Hispanic White teachers providing significantly lower scores compared to both non-Hispanic African American (d = -0.24, CI [-0.45, -0.03]) and Hispanic teachers (d = -0.33, CI [-0.49, -0.16]). Scores provided by non-Hispanic African American and Hispanic teachers did not differ significantly. No significant interactions or main effects were found for the IA scores.

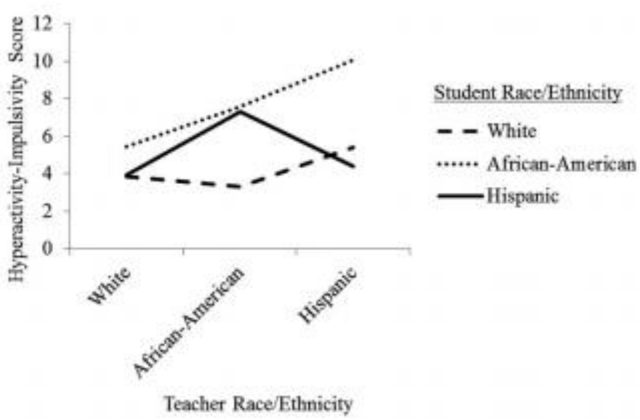

Figure 3. Means plot: student race-ethnicity and teacher race-ethnicity.

Table 7. Normative Data for Parent Ratings

		Inattention						Hyperactivity–impulsivity						Total score					
				Percentile						Percentile						Percentile			
Age	*n*	*M*	*SD*	80th	90th	93rd	98th	*M*	*SD*	80th	90th	93rd	98th	*M*	*SD*	80th	90th	93rd	98th
Males																			
5–7	244	6.14	5.79	9.0	15.0	17.0	23.0	5.78	5.35	9.0	15.0	17.0	18.9	11.92	10.31	18.0	27.0	31.0	40.0
8–10	238	5.74	5.40	9.0	12.0	16.0	21.4	4.48	5.58	8.0	13.0	16.0	20.0	10.21	10.20	16.0	25.1	30.0	42.4
11–13	245	6.41	6.13	11.0	15.0	17.0	25.1	3.63	4.73	7.0	10.0	11.5	19.0	10.05	9.82	18.0	23.0	25.5	38.0
14–16	230	5.60	6.08	11.0	15.0	18.0	22.6	2.47	3.55	4.0	6.0	8.0	14.1	8.07	8.96	14.5	23.0	26.0	36.4
17	101	4.99	5.57	9.0	12.9	17.0	23.3	2.01	3.46	3.0	8.0	9.0	13.9	7.00	8.16	11.0	18.5	20.0	35.4
Females																			
5–7	232	4.62	5.22	8.0	12.0	13.0	20.8	4.24	4.98	7.0	11.0	13.0	19.7	8.84	9.42	15.0	20.8	24.5	42.3
8–10	226	4.79	5.24	9.0	12.0	13.9	20.7	3.38	4.23	6.0	8.0	9.0	16.9	8.17	8.65	14.0	18.0	20.5	36.0
11–13	233	4.65	5.26	9.0	12.1	13.5	20.0	2.80	4.02	5.0	8.0	11.0	15.0	7.45	8.28	13.0	20.0	23.0	33.2
14–16	237	5.49	6.06	10.0	15.0	18.0	23.5	2.31	3.55	4.0	6.0	8.0	14.1	7.80	8.90	15.0	20.0	23.0	34.7
17	83	4.70	5.19	8.9	12.0	13.0	21.0	2.19	3.76	3.0	8.0	8.0	18.3	6.89	8.43	11.0	20.0	20.0	36.2

Normative Data

Normative data were computed separately for weighted parent and teacher ratings and presented for eight Gender Age (5- to 7-year-olds, 8- to 10-year-olds, 11- to 13-year-olds, 14- to 16-year olds, and 17-year-olds) groups. Because the DSM–5 diagnostic thresholds are different for 17-year-olds (i.e., requiring only five IA and/or hyperactive–impulsive symptoms), this group was reported separately. It should be noted, however, that the results of factorial invariance tests comparing 14- to 16-year-olds and 17- year-olds indicated that the measurement properties of the scores were substantively equivalent between age groups—parents: $\Delta\chi 2(50)$= 40.41, p $\succ$.05, ΔCFI = .001; teachers: $\Delta\chi 2(50)$= 52.19, p $\succ$.05, ΔCFI = .000. Additionally, t tests and effect sizes were computed to assess mean differences between the two age groups (14- to 16-year-olds vs. 17-year-olds). Mean scores did not differ significantly for IA or HI, with Cohen's d effect sizes ranging from 0.03 to 0.12, indicating that differences were small to negligible. Tables 7 and 8 list the means, standard deviations, and percentile scores for the IA, HI, and total scores for parent and teacher ratings, respectively. Four cutoff points are presented: 80th, 90th, 93rd, and 98th percentiles. These cutoff points can be used for screening risk (80th and 90th percentiles) and identification purposes (93rd and 98th percentiles; Power et al., 1998).

Discussion

This study supports the validity of the two-factor model of ADHD described in the DSM–IV–TR and DSM–5. As expected, the two-factor structure was supported for both parent and teacher ratings (e.g., Bauermeister et al., 1995; DuPaul et al., 1997, 1998). Also, like other studies, this investigation demonstrated that a factor solution that differentiates hyperactivity and impulsivity symptoms provided an acceptable fit to the data for parent and teacher ratings (e.g., Burns, Boe, Walsh, Sommers-Flanagan, & Teegarden, 2001; Ullebø et al., 2012). However, the very high correlation between the hyperactivity and impulsivity factors, consistent with previous research (Burns et al., 2001; Ullebø et al., 2012), raises serious questions about the uniqueness of these factors and the clinical utility of interpreting the hyperactivity and impulsivity dimensions separately. Further, the low number of items on the Impulsivity factor raises questions about whether this dimension assesses the breadth of this construct.

Table 8 Normative Data for Teacher Ratings

		Inattention						Hyperactivity–impulsivity						Total score					
				Percentile						Percentile						Percentile			
Age	*n*	*M*	*SD*	80th	90th	93rd	98th	*M*	*SD*	80th	90th	93rd	98th	*M*	*SD*	80th	90th	93rd	98th
Males																			
5–7	238	9.03	7.87	16.8	22.0	23.5	26.0	7.10	7.02	14.0	18.0	18.2	26.0	16.12	13.42	29.0	36.6	39.7	47.9
8–10	253	9.00	8.20	17.0	20.8	24.0	27.0	6.78	7.22	14.0	17.0	19.0	25.2	15.78	14.57	30.0	35.8	39.1	52.0
11–13	264	7.73	7.69	14.0	20.0	23.0	26.0	4.47	5.90	9.0	15.0	16.0	20.9	12.20	12.47	22.0	32.0	35.4	45.6
14–16	224	7.23	7.17	13.0	18.0	21.0	27.0	4.09	5.90	8.0	13.0	14.0	24.2	11.32	12.25	20.0	27.0	33.0	50.5
17	118	6.81	6.29	13.0	16.4	17.0	21.0	3.15	4.42	7.4	10.0	10.0	17.6	9.97	9.30	19.0	22.9	25.0	30.7
Females																			
5–7	228	6.09	6.93	13.0	17.0	18.0	23.0	4.84	6.11	9.0	14.0	17.0	22.0	10.93	12.29	23.3	30.0	33.0	42.1
8–10	244	5.33	6.54	10.0	17.0	18.0	23.9	3.06	4.71	6.0	11.0	13.0	16.9	8.39	10.43	17.0	24.0	28.6	39.2
11–13	247	5.63	6.85	10.0	17.0	18.3	25.0	2.68	4.24	5.0	9.0	9.7	15.0	8.31	9.98	17.0	24.0	27.0	33.1
14–16	222	3.14	4.69	7.0	10.0	11.0	17.4	1.65	3.31	2.0	5.2	8.0	12.0	4.79	7.46	9.0	14.0	18.0	30.2
17	103	4.14	5.40	8.0	11.0	13.0	24.2	1.26	2.83	2.0	4.0	6.9	12.0	5.40	7.16	10.0	18.0	19.4	27.0

This study is unique in that it demonstrates the invariance of the two-factor measurement model of ADHD across child gender, child race– ethnicity, child age groups, informant (parent vs. teacher), informant gender, and English–Spanish versions. To our knowledge, the invariance of the two-factor model across child gender, race– ethnicity, and age groups has not been previously examined. Demonstrating invariance as a function of child gender provides partial validation of the appropriateness of the ADHD Rating Scale–5 for both boys and girls. Demonstrating invariance across age groups is noteworthy given that the DSM–5 suggests alternative item wording for adolescents and that the ADHD Rating Scale–5 incorporates these recommendations into the adolescent version of the parent and teacher scales. Although prior research strongly suggests that the two-factor model is invariant across informants, this study is the first to demonstrate invariance as a function of informant gender. Overall, the findings regarding the invariance of the two-factor model support the practice of using the ADHD Rating Scale–5 across a wide range of child and informant groups.

Although a major goal of this study was to examine the structural validity of the two-factor, theoretical model delineated in the DSM–5, for purposes of comparison, a bifactor solution was also tested. In the bifactor model, explained common variance and omega hierarchical indicated that little reliable variance existed beyond that due to the General ADHD factor. In particular, the HI items in the bifactor model formed a weak group factor. As such, “the interpretation of the two subscales as precise indicators of unique constructs is extremely limited” (Reise, 2012, p. 691). Some researchers have suggested that bifactor models may be preferable to a simple two-factor structure (Martel, von Eye, & Nigg, 2010; Ullebø et al., 2012); however, several studies, including this investigation, have demonstrated that a bifactor version of the two-factor structure identifies a strong General ADHD factor and a distinct IA factor but a weak HI factor for both parents and teachers (e.g., Burns et al., 2014; Normand et al., 2012; Toplak et al., 2012; Ullebø et al., 2012). Future research is needed to provide an in-depth investigation of alternative models with regard to their empirical fit and theoretical meaningfulness based upon existing research on the genetics, neurobiology, and developmental course of this disorder.

Similar to other investigations (e.g., DuPaul et al., 1997, 1998; Reynolds & Kamphaus, 2004), this study confirmed child gender and age group differences on the ADHD dimensions. Although there was evidence of gender differences for parent ratings of the ADHD factors, differences were much more striking (medium effect sizes) for teacher ratings. With regard to age, on the HI factor, older children scored significantly lower than younger children; on the IA factor, age differences were nonsignificant on parent ratings and significant with small effects on teacher ratings. Taken together, such findings provide justification for presenting normative tables that incorporate child age and gender subgroupings in order to more accurately capture the degree to which ADHD symptoms deviate from developmental expectations, as required in the DSM–5.

This study also confirmed that ratings of ADHD symptoms differ as a function of child race– ethnicity. The pattern of findings differed by informant. Teachers rated non-Hispanic African American children higher than non-Hispanic White, Asian, and Hispanic children for IA and HI (small to medium effects). These findings are consistent with those derived using the previous version of the ADHD Rating Scale (DuPaul et al., 1997). Research is needed to understand factors that account for higher teacher ratings for non-Hispanic African American children.

In contrast to the findings obtained on teacher rating scales, there were significant differences in parent ratings of ADHD symptoms between Hispanic and non-Hispanic children. On the total ADHD scale, non-Hispanic African American and White children were rated higher than Hispanic children (small effects). On the IA factor, nonHispanic White children were rated higher than Hispanic children, and on the HI factor, non-Hispanic African American children were rated higher than Hispanic children (both small effects). These results are consistent with recent findings that rates of parent-reported ADHD are higher for non-Hispanic than Hispanic children (Visser et al., 2014). Additional research is needed to explicate the meaning of these findings. Research suggests that perceptions of the severity of child behavior may be influenced by a wide range of family, cultural, and socioeconomic factors, including level of family stress, singleparent status, number of children in the home, level of acculturation, and culturally influenced thresholds for differentiating normal from abnormal behavior (Cauce et al., 2002; Eiraldi, Mazzuca, Clarke, & Power, 2006).

The study also yielded findings pertaining to cultural factors influencing teacher ratings of ADHD symptoms. Teacher ratings of IA demonstrated substantial consistency; there were no main effects of teacher race–ethnicity and no student–teacher interaction effects of race–ethnicity. This finding is consistent with research demonstrating the similarity of teacher ratings and direct observations of classroom behavior for children of diverse racial groups (Epstein et al., 2005). With regard to ratings of HI, most of the interaction effects examined were not statistically significant, suggesting that teachers of varying racial–ethnic backgrounds generally rated students of diverse backgrounds similarly. An exception to this rule is that African American teachers rated Hispanic students significantly higher than nonHispanic White students (large effect size). Further, the findings uncovered a tendency for non-Hispanic White teachers to rate HI lower than Hispanic and non-Hispanic African American teachers regardless of the child's racial–ethnic group, although effect sizes were small. The tendency for Hispanic teachers to assign higher ratings of HI than White teachers is consistent with the findings of deRamirez and Shapiro (2005), although the latter investigators found this difference only when teachers rated Hispanic students. Replication of these findings is needed, especially given that some cell sizes in the analyses of interaction effects were relatively small.

The diagnostic assessment of ADHD ought to include an examination of symptom-related impairment in addition to an evaluation of ADHD symptoms. The ADHD Rating Scale–5 was designed to assess both symptoms and symptom-related impairments (DuPaul, Reid, Anastopoulos, & Power, 2014). This study investigated informant ratings of symptoms only; examining ADHDrelated impairments was beyond the scope of this investigation. Research is needed to examine how ratings of symptoms and impairments across informants can be combined in the diagnostic assessment of this disorder.

A limitation of the teacher sample in this study is that the response rate for those who were "off panel" and therefore less likely to represent national census targets was low (11.1%) in comparison to teachers who were "on panel" and likely to represent census targets (67.5%). To address this limitation, off-panel teachers were over-recruited, and only those who qualified based upon a consideration of census targets linked to demographic variables (i.e., student grade, race, ethnicity, geographic region) were included.

Due to the recruitment strategies used in the study, the teacher sample generally was representative of national census targets (see Table 3). With regard to ethnicity, the teacher sample closely approximates national estimates of the Hispanic population of children (about 23%). With respect to race, the sample closely approximates national estimates of the non-Hispanic White, Black, Asian, and Native American populations. A limitation of the teacher sample, however, is that students from the Northeast and Midwest were slightly overrepresented (by approximately 5%) and students from the South were underrepresented (by approximately 10%).

The parent sample generally reflected demographic characteristics of the U.S. population (see Table 2). Concerted efforts were made to include Spanish-speaking families and those with and without Internet access. Nonetheless, sampling methods may have resulted in an underrepresentation of families who are highly mobile, those who do not speak English or Spanish, and those with the lowest socioeconomic status. For example, in order to be included in the sampling panel, parents had to be able to obtain postal mail or have a telephone number that matched their mailing address.

For the teacher sample, nesting of students in a classroom was a potential limitation because the same teacher rated two students. Nonindependence of the teacher ratings could reduce standard errors of measurement, leading to an inflated Type I error rate. However, nesting is only problematic when design effects exceed a value of 2 (Muthén & Satorra, 1995). Because all of the clusters were small (two students per teacher), the resultant design effects were practically negligible ($\leqslant$1.5) and were not likely to result in an inflation of the Type I error rate.

In conclusion, the current findings provide support for the construct validity of the ADHD Rating Scale–5; its factor structure is compatible with the manner in which the DSM–5 conceptualizes ADHD. To address cross-informant differences that commonly occur as a result of the situational variability of ADHD symptoms (Barkley, 2015), the ADHD Rating Scale–5 provides separate normative tables for parents and teachers, which were derived from a large, nationally representative sample. Within each of these informant categories, the normative tables are further refined on the basis of empirically determined age and gender considerations. In so doing, the ADHD Rating Scale–5 normative tables afford researchers and clinicians alike a more accurate appraisal of the degree to which ADHD symptoms are developmentally deviant, a requirement for diagnosing ADHD in the DSM–5. Although race–ethnicity differences were detected, these were not incorporated into the normative tables for a variety of reasons, not the least of which is that these differences were not consistent across parent and teacher ratings. Until the meaning of such differences is clarified, researchers and clinicians would be well advised to use caution when interpreting findings obtained from racially and ethnically diverse populations.

References

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author.

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.). Washington, DC: Author.

Anastopoulos, A. D., & Shelton, T. L. (2001). Assessing attention-deficit/ hyperactivity disorder. New York, NY: Kluwer Academic/Plenum Press.

Barkley, R. A. (Ed.). (2015). Attention-deficit/hyperactivity disorder: A handbook for diagnosis and treatment (4th ed.). New York, NY: Guilford Press.

Bauermeister, J. J., Bird, H. R., Canino, G., Rubio-Stipec, M., Bravo, M., & Alegria, M. (1995). Dimensions of attention deficit hyperactivity disorder: Findings from teacher and parent reports in a community sample. Journal of Clinical Child Psychology, 24, 264 –271. http://dx .doi.org/10.1207/s15374424jccp2403_3

Baumgaertel, A., Wolraich, M. L., & Dietrich, M. (1995). Comparison of diagnostic criteria for attention deficit disorders in a German elementary school sample. Journal of the American Academy of Child & Adolescent Psychiatry, 34, 629 – 638. http://dx.doi.org/10.1097/00004583- 199505000-00015

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238 –246. http://dx.doi.org/10.1037/0033-2909 .107.2.238

Boomsma, A. (2000). Reporting analyses of covariance structures. Structural Equation Modeling, 7, 461– 483. http://dx.doi.org/10.1207/ S15328007SEM0703_6

Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). On the factor structure of the Beck Depression Inventory–II: G is the key. Psychological Assessment, 25, 136 –145. http://dx.doi.org/10.1037/a0029228

Browne, M. W., & Cudeck, R. (1993). Alternative ways of accessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136 –162). Newbury Park, CA: Sage.

Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? Journal of Abnormal Child Psychology, 29, 339 –349. http:// dx.doi.org/10.1023/A:1010314030025

Burns, G. L., de Moura, M. A., Beauchaine, T. P., & McBurnett, K. (2014). Bifactor latent structure of ADHD/ODD symptoms: Predictions of dual pathway/trait-impulsivity etiological models of ADHD. Journal of Child Psychology and Psychiatry, 55, 393– 401. http://dx.doi.org/10.1111/jcpp .12165

Cauce, A. M., Domenech-Rodríguez, M., Paradise, M., Cochran, B. N., Shea, J. M., Srebnik, D., & Baydar, N. (2002). Cultural and contextual influences in mental health help seeking: A focus on ethnic minority youth. Journal of Consulting and Clinical Psychology, 70, 44 – 55. http://dx.doi.org/10.1037/0022-006X.70.1.44

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Conners, C. K. (2008). Conners (3rd ed.). Toronto, ON, Canada: MultiHealth Systems. DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. International Journal of Testing, 13, 354 –378. http://dx.doi.org/ 10.1080/15305058.2013.799067

de Ramírez, R. D., & Shapiro, E. S. (2005). Effects of student ethnicity on judgments of ADHD symptoms among Hispanic and White teachers. School Psychology Quarterly, 20, 268 – 287. http://dx.doi.org/10.1521/ scpq.2005.20.3.268

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M., & McGoey, K. (1998). Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. Journal of Psychopathology and Behavioral Assessment, 20, 83–102. http://dx.doi .org/10.1023/A:1023087410712

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

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, 436 – 444. http://dx.doi.org/10.1037/1040-3590.9.4.436

DuPaul, G. J., Reid, R., Anastopoulos, A. D., & Power, T. J. (2014). Assessing ADHD symptomatic behaviors and functional impairment in school settings: Impact of student and teacher characteristics. School Psychology Quarterly, 29, 409 – 421. http://dx.doi.org/10.1037/ spq0000095

DuPaul, G. J., & Stoner, G. (2014). ADHD in the schools: Assessment and intervention strategies (3rd ed.). New York, NY: Guilford Press.

Eiraldi, R. B., Mazzuca, L. B., Clarke, A. T., & Power, T. J. (2006). Service utilization among ethnic minority children with ADHD: A model of help-seeking behavior. Administration and Policy in Mental Health and Mental Health Services Research, 33, 607– 622. http://dx .doi.org/10.1007/s10488-006-0063-1

Epstein, J. N., Willoughby, M., Valencia, E. Y., Tonev, S. T., Abikoff, H. B., Arnold, L. E., & Hinshaw, S. P. (2005). The role of children's ethnicity in the relationship between teacher ratings of attention-deficit/ hyperactivity disorder and observed classroom behavior. Journal of Consulting and Clinical Psychology, 73, 424 – 434. http://dx.doi.org/ 10.1037/0022-006X.73.3.424

Holland, M. L., Gimpel, G. A., & Merrell, K. W. (1998). Innovations in assessing ADHD: Development, psychometric properties, and factor structure of the ADHD Symptoms Rating Scale (ADHD-SRS). Journal of Psychopathology and Behavioral Assessment, 20, 307–332. http://dx .doi.org/10.1023/A:1021915606301

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

Martel, M. M., von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: Is there a "g" factor? Journal of Child Psychology and Psychiatry, 51, 905–914. http://dx.doi.org/10.1111/j.1469-7610 .2010.02232.x

McDonald, R. P. (1999). Test theory: A unified treatment. Mahwah, NJ: Erlbaum. Millsap, R. E. (2011). Statistical approaches to measurement invariance. New York, NY: Routledge/Taylor & Francis Group.

Molina, B. S. G., Smith, B. H., & Pelham, W. E. (2001). Factor structure and criterion validity of secondary school teacher ratings of ADHD and ODD. Journal of Abnormal Child Psychology, 29, 71– 82. http://dx.doi .org/10.1023/A:1005203629968

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), Sociological methodology (pp. 267–316). Washington, DC: American Sociological Association. http://dx.doi.org/10.2307/271070

Muthén, L. K., & Muthén, B. O. (1998 –2014). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén. Normand, S., Flora, D. B., Toplak, M. E., & Tannock, R. (2012). Evidence for a general ADHD factor from a longitudinal general school population study.

Journal of Abnormal Child Psychology, 40, 555–567. http://dx.doi.org/10.1007/s10802-011-9584-5

Pelham, W. E., Jr., Fabiano, G. A., & Massetti, G. M. (2005). Evidence based assessment of attention deficit hyperactivity disorder in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34, 449 – 476. http://dx.doi.org/10.1207/s15374424jccp3403_5

Pillow, D. R., Pelham, W. E., Jr., 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, 293–309. http://dx.doi .org/10.1023/A:1022658618368

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. The American Journal of Psychiatry, 164, 942–948. http://dx.doi.org/10.1176/ajp.2007.164.6.942

Power, T. J. (1992). Contextual factors in vigilance testing of children with ADHD. Journal of Abnormal Child Psychology, 20, 579 –593. http://dx .doi.org/10.1007/BF00911242

Power, T. J., Doherty, B. J., Panichelli-Mindel, S. M., Karustis, J. L., Eiraldi, R. B., Anastopoulos, A. D., & DuPaul, G. J. (1998). Integrating parent and teacher reports in the diagnostic assessment of ADHD. Journal of Psychopathology and Behavioral Assessment, 20, 57– 81. http://dx.doi.org/10.1023/A:1023035426642

Reise, S. P. (2012). The rediscovery of bifactor measurement models. Multivariate Behavioral Research, 47, 667– 696. http://dx.doi.org/ 10.1080/00273171.2012.715555

Reynolds, R. C., & Kamphaus, W. R. (2004). Behavior Assessment System for Children: Manual (2nd ed.). Circle Pines, MN: AGS Publishing.

Steiger, J. H., & Lind, J. C. (1980, May). Statistically-based tests for the number of common factors. Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, IA.

Swanson, J., Nolan, W., & Pelham, W. E. (1992). The SNAP-IV Rating Scale. Retrieved from http://www.adhd.net

Toplak, M. E., Sorge, G. B., Flora, D. B., Chen, W., Banaschewski, T., Buitelaar, J.,... Faraone, S. V. (2012). The hierarchical factor model of ADHD: Invariant across age and national groupings? Journal of Child Psychology and Psychiatry, 53, 292–303. http://dx.doi.org/10.1111/j .1469-7610.2011.02500.x

Ullebø, A. K., Breivik, K., Gillberg, C., Lundervold, A. J., & Posserud, M. B. (2012). The factor structure of ADHD in a general population of primary school children. Journal of Child Psychology and Psychiatry, 53, 927–936. http://dx.doi.org/10.1111/j.1469-7610.2012.02549.x

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M.,... Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention deficit/hyperactivity disorder: United States, 2003–

2011. Journal of the American Academy of Child & Adolescent Psychiatry, 53, 34 – 46. e2. http://dx.doi.org/10.1016/j.jaac.2013.09.001

Wainer, H. (1976). Estimating coefficients in linear models: It don’t make no nevermind. Psychological Bulletin, 83, 213–217. http://dx.doi.org/ 10.1037/0033-2909.83.2.213

Wolraich, M. L., Hannah, J. N., Baumgaertel, A., & Feurer, I. D. (1998). Examination of DSM-IV criteria for attention deficit/hyperactivity disorder in a county-wide sample. Journal of Developmental and Behavioral Pediatrics, 19, 162–168. http://dx.doi.org/10.1097/00004703- 199806000-00003

This article was published Online First May 25, 2015.

George J. DuPaul, Department of Education and Human Services, Lehigh University; Robert Reid, Department of Special Education and Communication Disorders, University of Nebraska–Lincoln; Arthur D. Anastopoulos, Department of Human Development and Family Studies, University of North Carolina at Greensboro; Matthew C. Lambert, Department of Special Education and Communication Disorders, University of Nebraska–Lincoln; Marley W. Watkins, Department of Educational Psychology, Baylor University; Thomas J. Power, Department of Pediatrics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, and Perelman School of Medicine, University of Pennsylvania.

Arthur D. Anastopoulos, George J. DuPaul, Thomas J. Power, and Robert Reid have a financial interest in the ADHD Rating Scale–5, which was used in this study to assess parent and teacher ratings of ADHD symptoms.

Correspondence concerning this article should be addressed to George J. DuPaul, Department of Education and Human Services, Lehigh University, 111 Research Drive, Bethlehem, PA 18015. E-mail: gjd3@lehigh.edu