Invariance of ADHD symptoms across sex and age: A latent analysis of ADHD and impairment ratings from early childhood into adolescence

Daniel R. Leopold,
University of Colorado Boulder

Micaela E. Christopher,
University of Colorado Boulder

Richard K. Olson,
University of Colorado Boulder

Stephen A. Petrill, and
Ohio State University

Erik G. Willcutt
University of Colorado Boulder

Abstract

A population-based longitudinal sample of 489 twin pairs was assessed at six time points over ten years to examine the measurement invariance and stability of attention-deficit/hyperactivity disorder (ADHD) symptoms, as well as the developmental relations between inattention (IN), hyperactivity-impulsivity (HI), and multiple aspects of functional impairment. Parent ratings of ADHD symptoms and functional impairment were obtained in preschool and after the completion of kindergarten, first, second, fourth, and ninth grades. Results of the temporal and sex invariance models indicated that parent ratings of the 18 ADHD symptoms function in the same manner for females and males from early childhood into adolescence. In addition to establishing this prerequisite condition for the interpretation of longitudinal and between-sex differences in the IN and HI symptom dimensions, cross-lagged models indicated that both IN and HI were associated with increased risk for both concurrent and future overall, social, and recreational impairment, whereas only IN was uniquely associated with later academic impairment. Taken together, the current results demonstrate that IN and HI are highly stable from preschool through ninth grade,

Point of Contact: Correspondence should be addressed to Daniel R. Leopold, Department of Psychology and Neuroscience, UCB 345, University of Colorado, Boulder, CO 80309-0345. Daniel.Leopold@colorado.edu; ORCID: 0000-0001-7250-9998.
Departmental affiliations: Daniel R. Leopold, Micaela E. Christopher, Richard K. Olson, and Erik G. Willcutt, Department of Psychology and Neuroscience, University of Colorado Boulder. Stephen A. Petrill, Department of Psychology, Ohio State University.

invariant between females and males, and indicative of risk for impairment in multiple areas, thereby providing strong support for the validity of the symptom dimensions among both sexes.

Keywords

inattention; hyperactivity; invariance; gender; impairment; longitudinal

In order to comprehensively evaluate the internal and external validity of attention-deficit/ hyperactivity disorder (ADHD) during the transition between the fourth and fifth editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV & DSM-5; American Psychiatric Association, 1994 & 2013), Willcutt and colleagues conducted a meta-analysis of 546 studies, including over 60,000 children and adolescents. Exploratory and confirmatory factor analyses of parent and teacher ratings indicated that inattention (IN) and hyperactivity-impulsivity (HI) are best conceptualized as distinct but highly correlated symptom dimensions (Willcutt et al., 2012). Although the optimal parameterization and nosology of ADHD continues to be debated, strong support exists for the measurement specification of the ADHD symptom dimensions using either two correlated factors (e.g., Willcutt et al., 2012) or a bi-factor model (Willoughby, Blanton, & Family Life Project Investigators, 2015), even across at least 15 countries (Bauermeister, Canino, Polanczyk, & Rohde, 2010).

Confidence in this dimensional view is limited by the fact that only a handful of studies have directly compared the IN and HI factor structure of ADHD symptoms in males and females (e.g., Burns, Walsh, Gomez, & Hafetz, 2006) or examined samples of children prior to the beginning of school (e.g., Friedman-Weieneth, Doctoroff, Harvey, & Goldstein, 2009), with Bauermeister et al. (2010) even reporting that the two-factor model was not upheld for preschool children. The paucity of studies means that questions remain regarding the measurement properties of these items across childhood, the IN and HI dimensions' differential trajectories and predictive validity, and the existence of item bias between males and females.

The current study addresses these open questions by analyzing parent ratings of DSM-IV ADHD symptoms and functional impairment collected at six time points over a 10-year period starting in preschool. In addition to replicating and extending results of earlier cross-sectional studies, these results provide new longitudinal results regarding the measurement qualities and longer-term risk associated with IN and HI. To set the stage for these analyses, we next summarize the existing literature on the developmental course of ADHD and associated functional impairment, highlighting several important gaps in knowledge that we are able to test directly using a longitudinal design and a large, population-based sample.

Developmental trajectories of IN and HI and risk for concurrent and future impairment

Results of cross-sectional studies indicate that both IN and HI symptoms are significantly associated with multiple domains of functional impairment (Willcutt et al., 2012), arguably

the most critical criterion for a mental disorder to be considered valid (e.g., Spitzer & Wakefield, 1999). Further, the discriminant validity of the IN and HI symptom dimensions is supported by studies that directly compared the functional correlates of the two dimensions. These studies suggest that IN is more closely associated with academic impairment, social withdrawal, and poor adaptive functioning, whereas HI has a stronger association with overt peer rejection, relational aggression, and frequency of accidental injuries (e.g., Lahey et al., 1998; Reynolds & Kamphaus, 2004; Willcutt et al., 2011, 2012).

Longitudinal studies have also reported that in comparison to children without ADHD, groups of children who met criteria for DSM-IV ADHD in childhood exhibited higher levels of global, social, and academic impairment five to nine years later (e.g., Hinshaw et al., 2006; Lahey & Willcutt, 2010; Owens, Hinshaw, Lee, & Lahey, 2009). However, less is known about the long-term functional outcomes associated with the IN and HI symptom dimensions. In one study that reported longitudinal results separately for IN and HI, Lahey and Willcutt (2010) found that levels of IN and HI in preschool predicted parent and teacher ratings of need for treatment and overall impairment nine years later, and levels of IN but not HI were associated with increased math difficulties and a greater likelihood that the child would be ignored by peers. Results of these longitudinal studies of clinic-referred samples of children suggest that IN and HI may be associated with different developmental outcomes, but the degree to which these distinctions generalize to unselected samples or the broader population is understudied.

Furthermore, although parent and teacher ratings of IN and HI have adequate test-retest reliability over periods less than one year (Willcutt et al., 2012), longitudinal studies of individuals who first received a clinical diagnosis of DSM-IV ADHD during childhood suggest that IN and HI symptoms may follow different developmental trajectories. Over the first nine years of a prospective longitudinal study, children first diagnosed with DSM-IV ADHD in preschool exhibited a significant age-related decline in HI behaviors that was not related to pharmacologic or psychosocial treatment, whereas levels of IN symptoms did not change significantly (Lahey et al., 2004; Lahey et al., 2005; Lahey et al., 1998; Lahey & Willcutt, 2010). A similar pattern was reported in a five-year follow-up study of a sample of females with DSM-IV ADHD who were first assessed between 6 and 12 years of age (e.g., Hinshaw, 2002; Hinshaw, Owens, Sami, & Fargeon, 2006).

Although these previous studies provide important support for the validity of the distinction between IN and HI symptoms, several key questions remain. Many key studies recruited participants through clinics, which could potentially overestimate the relation between ADHD symptoms and functional impairment due to the extensive and severe range of difficulties often observed in clinical samples. In addition, only a handful of clinical studies have examined the extent to which early IN and HI are associated with increased risk for negative outcomes later in development (e.g., Lahey, Pelham, Loney, Lee, & Willcutt, 2005; Lahey et al., 1998; Lahey & Willcutt, 2010), and even fewer studies have systematically tested for the potential influence of demographic variables such as sex (e.g., Burns et al., 2006).

Sex differences in ADHD symptoms

Meta-analyses of population-based samples indicate that males are 2 – 3 times more likely than females to meet DSM-IV diagnostic criteria for ADHD (Willcutt, 2012). Similarly, males exhibit higher mean levels of IN and HI than females in both selected and unselected samples (e.g., Burns et al., 2006; Gaub & Carlson, 1997; Gershon, 2002). However, the reasons for these sex differences remains largely unknown.

One possible explanation is that the symptoms of ADHD might be less internally valid for females than males, which may then be reflected in lower reliability and a less robust factor structure of ADHD symptoms in females. Alternatively, ADHD might be identified more frequently in males due to a stronger association with specific aspects of impairment or concurrent psychopathology that lead parents and teachers to more frequently endorse ADHD symptoms. Consistent with this hypothesis, two meta-analyses found that males with ADHD exhibited slightly higher levels of externalizing behaviors than females, whereas females were more likely to exhibit internalizing symptoms (Gaub & Carlson, 1997; Gershon, 2002). However, effect sizes were small for both of these comparisons ($d = .1 – .2$), and the meta-analyses reported no sex differences in the small number of studies that compared males and females with ADHD on measures of other psychopathology or academic, social, or neuropsychological functioning. Both reviews concluded by calling for additional research to better understand the etiology of the robust sex difference in the prevalence of ADHD, and highlighted a particular need for more studies of population-based samples to avoid potential biases due to differential rates of referral for clinical services in males and females.

The current study

This study was designed to clarify the measurement properties and developmental course of IN, HI, and associated functional outcomes over a ten-year period between preschool and the end of ninth grade. Ratings of IN, HI, and functional impairment were collected six times during this period from the parents of an unselected population-based sample of 489 twin pairs (i.e., 978 children). Structural equation modeling (SEM) was utilized to address three questions regarding the measurement properties, developmental trajectory, and risks associated with ADHD:

1. Is the factor structure of ADHD stable across development?

Initial confirmatory factor analyses tested whether parent ratings of ADHD symptoms fit the bi-dimensional DSM-IV model at each assessment point. We then examined temporal measurement invariance was examined across the full ten-year period. We predicted that the DSM-IV model with correlated dimensions of IN and HI symptoms would provide a good fit to the data at each time point as well as showing temporal measurement invariance across the six assessment points. Although we expected the population mean of HI ratings to decline more than the mean of IN ratings, we anticipated that the rank order of individuals would be consistent over time, leading to high IN and HI stability correlations across time.

2. Does the factor structure or developmental course of ADHD differ between males and females?

Expanded multiple group SEMs were fit with sex as a grouping variable to test whether the factor structure of ADHD symptoms and the overall pattern of results could be equated in males and females. Though most studies have compared males and females on clinical correlates of groups defined with ADHD, few have factor analyzed sex differences in ADHD at either a single time point or across development. Consistent with these previous reports of relatively few sex differences in analyses of total ADHD symptoms, we hypothesized that the measurement of ADHD would be invariant across sex, despite elevated ratings of males' behavior.

3. Are IN and HI symptoms associated with concurrent and future risk for negative functional outcomes?

Following an initial series of cross-sectional analyses on the relations between IN, HI, and different aspects of functional impairment, latent cross-lagged models provided a novel and stringent test of whether earlier levels of IN and HI uniquely predicted subsequent impairment. Although we expected both dimensions to be associated with all aspects of functional impairment, we predicted that IN symptoms would be more strongly associated with academic difficulties, whereas elevations of HI would be more strongly associated with early social impairment.

Method

Participants

The participants in the present study were 978 individuals drawn from 224 monozygotic (MZ; i.e., identical) and 265 dizygotic (DZ; i.e., fraternal) same-sex twin pairs first assessed during the summer prior to starting kindergarten ($N_{pairs} = 482$, $M_{age} = 4.9$ years, $SD_{age} = 0.2$ years; 49.7% male; note: the term *sex* is used instead of *gender* throughout the manuscript, as parents were initially asked to indicate the twins' biological sex). For all analyses, one randomly selected member from each pair was used to control for the non-independence of individuals within each twin pair. All participants were part of the Colorado component of the International Longitudinal Twin Study of Early Reading Development (ILTSERD; e.g., (Byrne et al., 2002; Christopher et al., 2013; Olson et al., 2011), and were recruited from the Colorado Twin Registry based on birth records. The Colorado Twin Registry includes information on over 90% of all twin births in Colorado, 60% of whom were able to be contacted, and comparisons with available normative data on several measures suggest that the current sample is representative of the overall population in the state (e.g., Christopher et al., 2015).

After the initial preschool assessment, participants were assessed again in the summers following kindergarten ($N_{pairs} = 453$, $M_{age} = 6.3$ years, $SD_{age} = 0.3$ years), first grade ($N_{pairs} = 442$, $M_{age} = 7.4$, $SD_{age} = 0.3$), second grade ($N_{pairs} = 451$, $M_{age} = 8.5$, $SD_{age} = 0.3$), fourth grade ($N_{pairs} = 445$, $M_{age} = 10.5$, $SD_{age} = 0.3$), and ninth grade ($N_{pairs} = 453$, $M_{age} = 15.5$, $SD_{age} = 0.3$). Retention was excellent from preschool through the end of ninth grade (92 – 94%).

Procedures

Overall testing procedures for the ILTSERD are described in detail in previous papers (e.g., Byrne et al., 2002; Christopher et al., 2013; Willcutt et al., 2007). Briefly, at each wave the twins completed a battery of measures related to reading development in an individual testing session while one parent or caregiver completed a battery of questionnaires that included the measures described in this report. All study procedures were fully approved by the Institutional Review Boards of the University of Colorado Boulder. Informed consent or assent was obtained from all participants and their parents at initial enrollment and at each follow-up assessment.

Measures

DSM-IV ADHD symptoms

The *Disruptive Behavior Rating Scale* (DBRS; Barkley & Murphy, 1998) was used to obtain parent ratings of the 18 symptoms of DSM-IV ADHD. The majority of ratings were completed by mothers at all time points (90 – 95%). On the DBRS, the parent is asked to indicate how often in the last 6 months each of the 18 DSM-IV ADHD symptoms is true on a 4-point Likert scale (0 = *never or rarely,* 1 = *sometimes,* 2 = *often,* and 3 = *very often*). This scale has been widely used for the assessment of ADHD symptoms, demonstrating strong internal and test-retest reliability among children and adolescents, as well as a robust literature of correlations with both internalizing and externalizing psychopathology. Cronbach's alphas for the IN and HI dimensions ranged from .89 to .93 and .86 to .88, respectively. All internal reliability and test-retest reliability estimates between the six assessment points can be found in Table 1.

Functional Impairment

Seven items measuring functional impairment (interferes with home life and family; interferes with child interactions; interferes with adult interaction; interferes with community activities; interferes with educational activities; interferes with recreational activities; and interferes with daily responsibilities) were also administered in the same format as the ADHD items described above, with slight changes to the 4-point Likert scale (0 = not at all; 1= just a little; 2 = quite a bit; 3 = very much). Parents were again asked to complete their ratings regarding the past six months. Cronbach's alphas for the functional impairment scale ranged from .86 to .90. Internal and test-retest reliability estimates can be found in Table 1.

Analytic Strategy

Advantages of a latent trait approach

While many studies have used correlations or linear regressions to assess the stability of ADHD and the relation between ADHD symptoms and a range of external correlates, SEM and confirmatory factor analyses have several unique advantages for analyses that are designed to investigate issues of temporal stability and predictive utility. By first exploring whether individual items have statistically equivalent factor loadings and intercepts across time, stronger claims can be made about the relevant constructs and the extent to which they

change between measurement occasions. Furthermore, by using unobserved (i.e., latent) constructs composed of the shared, reliable variance of measured items or symptoms rather than the simple sum or mean of related items, both internal and external relationships can be explored with reduced measurement error.

Structural equation models

Structural and measurement model analyses were conducted using the Mplus statistical software package (Version 7.4; Muthen & Muthen, 2012). For item-level analyses, items were treated as ordered categorical manifest variables using the robust weighted least squares estimator (WLSMV). For analyses using parcels, parcels were treated as approximately continuous manifest variables using the robust maximum likelihood estimator (MLR). Robust estimation was used for both types of analyses in order to adjust for the non-normality that is characteristic of symptom ratings data (i.e., positively skewed and leptokurtotic). Model fit was assessed with the robust comparative fit index (CFI; study criteria of at least .90, with .95 being ideal), the Tucker-Lewis Index (TLI; study criteria .90), and the robust root-mean-square error of appoximation (RMSEA; study criteria . 08).

Parcels as manifest variables—An additional benefit of SEM is the use of parcels, whereby individual items are combined (e.g., taking the mean of multiple items) and the new composites are used as manifest variables. Parceling reduces the amount of unreliable (i.e., error) and item-specific variance and decreases the likelihood of Type II errors (Little, 2013; Little, Rhemtulla, Gibson, & Schoemann, 2013). Item-level analyses were first used to establish like-item loading (weak) and threshold (strong) invariance across time points. If weak and strong invariance was upheld, parcels were used to simplify estimation of longitudinal cross-lag models.

The ADHD-IN and ADHD-HI factors were each defined by three parcels of three items per parcel. Following the procedures recommended by Little (2013), unstandardized loadings from the strong invariant model were used to assign items to a parcel that would maximize the likelihood of homogenous parcels (see also Burns, Servera, Bernad, Carrillo, & Geiser, 2014). Specifically, the items with the highest and lowest unstandardized loadings were assigned to parcel 1, followed by the next highest and next lowest items to parcel 2, and so on until all items were assigned to a parcel. IN Parcel 1 involved the *attention to details, follow through*, and *easily distracted* symptoms; IN Parcel 2 the *sustaining attention, organization*, and *mental effort* symptoms; IN Parcel 3 the *does not listen, loses things*, and *forgetful* symptoms. HI Parcel 1 involved the *fidgets/squirms, runs/climbs*, and *talks excessively* symptoms; HI Parcel 2 the *leaves seat, blurts answers*, and *awaiting turn* symptoms; and HI Parcel 3 the *playing quietly, on the go/driven*, and *interrupts/intrudes* symptoms. Overall Impairment Parcel 1 consisted of the *interferes with child interactions* and *interferes with daily responsibilities* items; Overall Impairment Parcel 2 the *interferes with adult interactions* and *interferes with educational activities* items; and Overall Impairment Parcel 3 the *interferes with home life and family, interferes with community activities*, and *interferes with recreational activities* items. Each parcel was then used as a

manifest variable. Thus, each latent construct (IN, HI, or Overall Impairment) was represented by three manifest variables.

Criteria and procedure for invariance tests—Because the chi-square difference test is known to detect small discrepancies of no theoretical or practical consequence in large samples (Chen, Sousa, & West, 2005), changes in CFI, TLI, and RMSEA were used to assess the invariance of model constraints at the level of the indicators (i.e., loadings and thresholds). Specifically, if the decrease in CFI was less than 0.01, and the TLI and RMSEA showed little change (Chen, 2007; Little, 2013), the imposed constraints were assumed to be invariant. In order to test for measurement invariance of the three constructs across time, we first evaluated the 25 items (i.e., nine IN, nine HI, and seven functional impairment items) by constraining like-item loadings and thresholds to be equal across the four time points covering the largest time frame with evenly spaced assessments (i.e., pre-K, post 1st, post 4th, and post 9th grade). The post kindergarten and post 2nd grade waves were excluded to simplify model estimation with manifest variables. These analyses included correlated residuals for like-items across all occasions. Using the same 25 manifest variables at the four occasions, multiple group tests were then carried out to determine if the individual items showed measurement invariance across sex.

Following the recommendation by Little (2013), multiple group models of configural, strong, and latent mean, variance, and covariance invariance were then tested between females and males using parceled indicators, as described above. For these analyses, chi-square difference tests using the scaling correction factor as outlined by Muthén and Muthén (2012) were used to evaluate the invariance of structural parameters. Given sample size considerations and the higher power associated with longitudinal SEM models, a more stringent p-value was used to determine significance of these comparisons ($p < .005$) (Little, 2013). For latent factor mean comparisons, an omnibus test constraining all factor means between groups was first performed, followed by tests individually releasing constraints on the HI, IN, and then functional impairment factors between males and females. These multiple group comparisons test whether the IN, HI, and functional impairment constructs are factorially invariant across sex. If the variance-covariance matrices can be equated for males and females, the entire sample can then be analyzed as a single group in subsequent analyses.

Longitudinal cross-lagged analyses—Upon establishing the temporal and sex measurement invariance of IN, HI, and overall functional impairment across all measurement occasions, the three latent simplex models (i.e., autoregressive models in which, for example, each latent IN factor is regressed onto the IN factor from the prior occasion) were estimated. A latent cross-lagged model was then estimated, combining the three latent simplex models into a single model and regressing each occasion's constructs onto those of the previous occasion. To examine separately the relations between ADHD symptoms and social, academic, and recreational impairment while maintaining an identified model, three additional three-level cross-lagged models were estimated in which each of these impairment domains was modeled as a single item or mean of items. The social impairment model included the *interferes with home life and family*, *interferes with child*

interactions, and *interferes with adult interactions* items; the academic impairment model included the *interferes with educational activities* item; and the recreational impairment model included the *interferes with recreational activities* item.

Missing data

Due to the high rate of retention across all waves of the study (>92%), data were missing for a relatively small proportion of observations at each wave (mean = 9%, range for individual items = 1 - 24%). Covariance coverage for parents' ratings ranged from 0.58 to 0.99. The percentage of missing data for each variable ranged from 1.4% to 29.0%, with a mean (standardized deviation) of 12.3% (8.6%). Analyses of individual items used the WLSMV estimator with a pairwise approach to missing data. Analyses using item parcels used the MLR estimator with a full information maximum likelihood approach to missing data.

Results

Descriptive information on manifest variables

Developmental changes in levels of IN, HI, and functional impairment—Table 1 provides the descriptive results for the manifest variables. The mean level of HI symptoms in the population declined significantly across development (Table 1), with medium to large effect sizes for paired t-tests of the differences between ratings obtained in preschool and kindergarten and ratings obtained after 4th and 9th grade (d = .4 - 1.0). In contrast, mean levels of IN and functional impairment generally remained stable (d < .2 for all changes between years). Mean ratings of IN, HI, and overall impairment were all higher for males than females (see Table 1; mean d for IN, HI, and impairment = .32, .35., and .27, respectively).

Table 1 also shows the bivariate correlations between the mean ratings of IN, HI, and functional impairment at each of the six assessment points. All correlations were significant and medium to large in magnitude for IN, HI, and overall functional impairment, and all results followed the expected longitudinal pattern in which ratings at adjacent time points correlate more highly with one another than with ratings at more distant time points. Results were similar when functional impairment was subdivided into measures of academic and social impairment, with slightly lower longitudinal correlations observed for the single-item measure of impairment during recreational activities (see online supplementary material, Supplemental Table 1).

The relation between IN and HI and risk for concurrent functional impairment—Bivariate correlations between manifest IN and HI were moderate at all time points (r = .52 - .59), and IN and HI were also significantly associated with all aspects of functional impairment at each assessment (Table 2). When measures of each aspect of functional impairment were regressed onto both symptom dimensions simultaneously, both IN and HI were independently associated with overall impairment, social impairment, and recreational impairment at all six assessments. In contrast, only IN was independently associated with academic impairment after the beginning of first grade (i.e., both dimensions were

independently associated with concurrent academic impairment prior to and after kindergarten).

Factor structure and reliability—Individual IN, HI, and functional impairment items loaded strongly on the corresponding latent trait at each of the assessment points (mean standardized loading = .80 for IN, .77 for HI, and .82 for impairment), providing strong initial support for the internal validity of the DSM-IV model. Similarly, estimates of internal consistency were high for composite measures of IN, HI, and each aspect of functional impairment at all time points (α = .86 – .93; Table 1).

Analyses using individual items

Temporal and sex invariance—Due to the different developmental trajectories of IN and HI, as well as mean differences between females and males, separate temporal and sex invariance models were fitted to test whether ratings of ADHD symptoms and functional impairment measured the same constructs across development and sexes. Table 3 summarizes the results of the invariance analyses for like-item loadings and thresholds across the six time points. The configural model demonstrated an acceptable and close fit to the data (CFI = .94; TLI = .94; RMSEA = .030), and the strong invariant model with loadings and thresholds constrained to be equal at all time points did not result in a significant decrease in fit. The baseline model for the multiple groups sex invariance tests also yielded an acceptable and close fit (CFI = .94; TLI = .94; RMSEA = .028), and the strong invariant model did not result in a meaningful decrement of fit. These results indicate that these constructs are consistent across both development *and* sex, and that any differences over time or between males and females are not due to changes in the constructs' measurement properties.

Analyses using parceled items

Sex invariance—Finally, a multiple group model with sex as a grouping variable demonstrated very good and close fit (Table 3; CFI = .96; TLI = .95; RMSEA = .037), and the strong invariant and latent variance-covariance invariant models did not result in significant decrement of fit. Therefore, although mean ratings on IN, HI, and functional impairment were higher for males than for females, these results indicate that the covariance structure is similar for both sexes. Table 4 presents the latent factor means and standard deviations for the full sample, as well as separately for females and males. The latent effect sizes for sex differences were generally small in magnitude for the IN, HI, and overall impairment factors ($\bar{d}$ = .32, .37, and .23, respectively). Thus, although latent mean invariance was not upheld, sex only accounted for 10%, 13% and 5% of the variance in IN, HI, and overall impairment factor means, respectively. For the remainder of the analyses reported in this paper, the entire sample was analyzed as a single group (to confirm that this did not bias the results, analyses were also completed separately in males and females, and the overall pattern of results was nearly identical to the analyses of the full sample).

Developmental stability of ratings of ADHD symptoms and functional impairment—Latent trait models were then used to examine simultaneously the overall stability of IN, HI, and functional impairment from preschool through ninth grade. Mean

standardized loadings of the parcels on the IN, HI, and functional impairment factors were high (IN $M = .89$; HI $M = .84$; and functional impairment $M = .86$), and the composite reliability coefficients (true-score variance) across occasions were excellent for all three factors (IN $M = .95$; HI $M = .95$; and functional impairment $M = .93$). These results confirmed that the models with parcels as manifest variables also displayed excellent psychometric properties. These parcels were then used to construct latent factors at each time point for the estimation of both simplex and cross-lagged models, maintaining the loading and intercept constraints of the strong invariant model.

The simplex models of ratings of IN, HI, and functional impairment from preschool through post 9th grade each displayed very good and close fit (CFIs >.97, TLIs >.96, and RMSEAs <.054; Supplemental Table 2). Furthermore, the standardized stability coefficients indicate significant temporal stability for both IN and HI ($\bar{r} = .70$ and $.74$, respectively). Stability coefficients were also significant but marginally lower for the functional impairment latent trait ($\bar{r} = .62$), with the strongest evidence of stability in mid to late childhood.

Longitudinal models—Finally, the developmental relationships between IN, HI, and functional impairment were examined by fitting a full, latent cross-lagged model with the IN, HI, and functional impairment parcels (Figure 1). The full cross-lagged model for IN, HI, and overall functional impairment provided a very good and close fit to the data ($X^2(1249) = 1875.3$, $p < .001$; CFI = .97; TLI = .96; RMSEA = .032 [.029–.035]. In order to simply this figure, the factor correlations and disturbance parameters can be found in Supplemental Table 3. The significant cross-lagged paths, which represent standardized partial regression coefficients, indicate cumulative increases in overall impairment if one is rated highly on IN after kindergarten, first, second, and fourth grades (i.e., increases in IN at these time points predict increases in impairment at the next measurement occasions above and beyond the predictive power of earlier increases in HI and impairment). Although only at one time point, this same result holds true if one is rated highly on HI after kindergarten. Of note, cumulative increases in IN were also evident if one is rated highly on overall impairment after first and fourth grades.

Similar cross-lagged models were then fit to mean values for ratings of social impairment, academic impairment, and impairment in recreational activities. Each of these models also displayed a very good and close fit (Supplemental Table 4). After accounting for previous social impairment, HI predicted new elevations of social impairment after kindergarten, first, and second grade, and IN predicted new elevations of social impairment after first, second, and fourth grades (Supplemental Figure 1). IN symptoms also predicted new academic impairment at each subsequent assessment, even over the five-year span from the summer after fourth grade to the end of ninth grade (Supplemental Figure 2). In contrast, HI symptoms were not uniquely associated with elevated risk for future academic impairment at any of the assessment points. Prior academic impairment, however, was significantly associated with IN elevations after kindergarten, first, and ninth grades. Finally, earlier HI symptoms predicted a unique increase in impairment in recreational activities at the first three outcome assessments (i.e., after kindergarten, first, and second grades), and prior elevations of IN symptoms predicted new recreational impairment elevations over the remaining years (i.e., at the end of fourth and ninth grades; Supplemental Figure 2).

Discussion

A ten-year longitudinal study of 489 individuals first assessed prior to kindergarten was used to examine the temporal and sex invariance of the 18 ADHD symptoms, as well as their associations with multiple aspects of functional impairment. To our knowledge, the results demonstrate the measurement invariance of the ADHD symptoms across both age and sex in a more comprehensive manner than any prior study. After summarizing the implications of the current results for developmental models of impairment related to ADHD, we discuss the results' consequences for diagnostic models of ADHD. We then describe several important limitations of the current study and highlight key directions for future research on ADHD symptom dimensions and their relations with different aspects of functional impairment.

Structure and stability of individual differences in ADHD symptom dimensions—Separate confirmatory factor analyses at each assessment point provided strong support for the distinction between the DSM-IV IN and HI symptom dimensions from preschool through early adolescence, and simultaneous SEM of ratings at all six time points indicated that parent ratings of IN and HI measure the same constructs across development. Further, high stability correlations for both IN and HI in the current longitudinal analyses add to a growing literature that suggests that the rank order of individuals in the population remains relatively stable for both IN and HI despite a decline in mean levels of HI across development.

The decline in HI symptoms as children get older is consistent with the results of earlier longitudinal studies of children with a clinical diagnosis of ADHD (Hinshaw et al., 2006; Lahey & Willcutt, 2010), indicating that this pattern is present in the overall population and is not restricted to individuals with ADHD. The different developmental trajectories of the symptom dimensions may also help explain why some individuals appear to shift systematically from the combined presentation of ADHD to the inattentive presentation over time (Willcutt et al., 2012).

Sex differences (and similarities)—The current results also replicate the well-established finding that males are rated higher on the IN and HI symptoms than females (e.g., Gaub & Carlson, 1997; Gershon, 2002; Willcutt, 2012). The consistency of this pattern across six assessment points in an unselected population-based sample provides additional confirmation that the higher rate of ADHD in males is not simply a clinical selection artifact or the result of item bias. Multiple group SEM also indicated that the factor structure of ADHD symptoms could be equated between both sexes. The finding of both measurement and structural invariance across the sexes is of arguably greater importance than mean differences in symptom ratings. The results demonstrate that the ADHD symptoms function in the same manner for both males and females. This important prerequisite not only provides support for the common practice of combining male and female samples for research purposes, but also indicates that the mean differences between sex are true differences in the ADHD construct (i.e., not due to item bias, differential item functioning, or changes in the rating scale's measurement properties over time). These results replicate prior sex invariance findings for the DSM-IV ADHD dimensions (Burns et al., 2006),

reproduce null results reported by earlier meta-analyses of sex differences in impairment (Gaub & Carlson, 1997; Gershon, 2002), and extend these findings by demonstrating that these results hold over a ten-year period from early childhood through early adolescence.

The current results thus provide important additional support for the validity of the IN and HI symptom dimensions in both males and females, but they do not explain the higher prevalence of ADHD in males beyond establishing that this difference is not due to differential item functioning or changes in the symptoms' measurement properties between sex or over time. Furthermore, despite the higher prevalence and mean ratings for males, sex only accounts for 10% of the variance in IN factor means and 13% for HI. One recent paper suggests that a small proportion of the difference in ADHD symptoms between males and females may be mediated by females' stronger performance on measures of processing speed, but this effect is relatively small and most of the sex difference remained unexplained (Arnett, Pennington, Willcutt, DeFries, & Olson, 2015). Further research is needed to understand sex differences in the prevalence of ADHD.

IN and HI as risk factors for concurrent and future negative outcomes—IN and HI were associated with multiple aspects of functional impairment in cross-sectional analyses at all time points, and latent cross-lagged models indicated that both symptom dimensions were associated with significant impairment at future assessments even after controlling for earlier levels of impairment. In contrast to elevated IN ratings' association with subsequent overall impairment at four of the five possible waves (i.e., after kindergarten, first, second, and fourth grades), only increases in HI ratings after kindergarten were uniquely associated with cumulative increased in later overall impairment. At the level of specific aspects of concurrent and future impairment, however, important differences emerged between the dimensions. Only IN was independently associated with concurrent and future academic difficulties, suggesting that the significant bivariate association between HI and academic impairment is explained by variance shared with IN rather than a unique association with HI per se. This pattern of results is highly consistent with earlier cross-sectional studies of various academic measures (see review by Willcutt et al., 2012) and underscores the unique importance of early IN symptoms as predictors of reading difficulties in early (Dittman, 2016) and middle (Pham, 2016) elementary school and critical targets for early identification and intervention. Furthermore, the significant cross-lagged paths from earlier overall impairment, and especially academic impairment, to IN may reflect the effects of early academic demands on IN behaviors (Brosco & Bona, 2017). These paths may also reflect the finding that, at least for a subset of individuals, academic impairment – and in particular reading difficulties – may either cause later attentional difficulties or be related to IN symptoms through shared genetic and cognitive risk factors (McGee, Prior, Williams, Smart, & Sanson, 2002; Willcutt, Betjemann, Pennington, Olson, DeFries, & Wadsworth, 2007).

In contrast to the results for academic functioning, both HI and IN were independently associated with social and recreational impairment in cross-sectional analyses at all six assessment points. Both IN and HI also predicted significant new impairment in these domains at several later assessments. However, consistent with recent meta-analytic results (Ros & Graziano, 2017), the effects were strongest for HI at the earlier assessments, whereas

levels of IN were only associated with future recreational impairment after fourth and ninth grade. Although the reasons for these differences are unclear, they may reflect the impact of early impulsivity on social and recreational functioning or changes in the nature of recreational activities later in development to include more significant attentional demands. Future studies that examine the relations between IN and HI and a wider range of measures of recreational and social functioning would provide a useful extension of the current results to test these and other possibilities.

Implications for diagnostic models of ADHD—The current results replicate and extend results from previous cross-sectional studies of the validity of the DSM-IV model of ADHD (see Willcutt et al., 2012). Importantly, results indicate that the symptoms of ADHD are invariant and function similarly between males and females and from early childhood through adolescence. As put forth in the DSM-IV, the IN and HI dimensions are correlated but clearly separable constructs that are independently associated with both concurrent impairment and increased risk for important negative outcomes later in development.

Although the current results do not have direct implications for decisions regarding the inclusion of nominal diagnostic subtypes as part of the ADHD diagnosis, the marked developmental instability of the DSM-IV subtypes and other results continue to call into question the utility of a subtype model to describe heterogeneity among individuals with ADHD (Lahey & Willcutt, 2010; Willcutt et al., 2012). Furthermore, the nominal subtypes offered little in the way of optimizing intervention efforts or clarifying causal pathways. The DSM-5 symptom presentations were thus intentionally framed to reflect potential developmental changes over time while retaining "subtype-like" categories. This use of presentations rather than subtypes appears to be consistent, for example, with the developmental declines in HI mentioned above.

As an alternative, the current results, as well as those from recent bi-factor analyses of ADHD symptoms in childhood and a recent study of the dimensionality of ADHD in adulthood (Hartung et al., 2016), provide additional empirical support for a model that would incorporate dimensional modifiers that reflect the number of IN and HI symptoms at the time of assessment (e.g., mild, moderate, and severe for 0–2, 3–5, and 6 or more current symptoms; see Willcutt et al., 2012). Such a model might also encourage research on the developmental course and biobehavioral underpinnings of these presentations. At a minimum, the current results argue strongly that separate IN and HI dimensions should be retained in the diagnostic criteria for ADHD. These results also underscore the impairing nature of IN across development and the clinically salient effects of HI, especially during the earliest educational years.

Limitations and future directions

A primary strength of the current study is the use of a large community sample that was first assessed prior to the beginning of school and then tested five times over a ten-year period ending after ninth grade. The same measures of ADHD and functional impairment were obtained at all assessments, and the low rate of attrition (approximately 92% retention through ninth grade) helped to minimize bias and maximize statistical power. Despite these

strengths, the current study also has several limitations that should be considered when interpreting the results.

Use of twins—All participants in the current study were members of same-sex twin pairs that were recruited through a community twin registry. While this sampling procedure facilitated the recruitment of a sample that is generally representative of the overall population of twins in Colorado, any unique effects of being a member of a twin pair may limit generalization to the larger population of singletons. Therefore, the current conclusions would be strengthened if they were replicated in a longitudinal sample of nontwins.

Measurement of ADHD—As part of the aims of the larger study, the measures of ADHD and functional impairment were added as one part of a battery of parent report questionnaires designed to assess a range of factors that might covary with individual differences in academic development. Due to time and budget constraints, DSM-IV ADHD was defined by parent ratings on the DBRS rather than a full structured diagnostic interview. In a previous study we found excellent agreement between parent ratings on the DBRS and a DSM-IV structured interview, suggesting that these methods are likely to yield similar results (Willcutt et al., 2010).

A second important measurement limitation is the fact that all ratings were completed by parents. Teacher ratings were not included in the current longitudinal analyses because they were only available after second grade. However, secondary analyses of the data collected after second grade indicated that correlations were moderate between parent and teacher ratings of IN ($r = .55$) and HI (r = .42), and the magnitudes of these correlations are similar to the pooled correlations reported in the recent meta-analysis of previous studies of DSM-IV ADHD (Willcutt et al., 2012). Further, the overall pattern of results when teacher ratings of functional impairment at the end of second grade were predicted by parent ratings of IN and HI was very similar to the results based only on parent ratings. Consistent with cross-setting invariance findings for the ADHD symptom dimensions (Burns et al., 2014; 2016), these results collectively suggest that the overall pattern of results are likely to have been similar if teacher ratings were included.

Nonetheless, the parent ratings available for all waves in the current study cannot rule out the potential influence of rater effects. Future studies that include both parent and teacher ratings and other functional impairment measures would provide a useful extension of the current results. By incorporating information from multiple raters, future studies could test the extent to which any observed within-trait stability or cross-trait prediction is due to rater-versus trait-specific variance.

Measures of functional impairment—The criteria for ADHD in DSM-IV and DSM-5 include detailed operational definitions of the nine symptoms of IN and HI. In contrast, little specific guidance is provided regarding the measurement of functional impairment. Although the seven impairment items in the current study were drawn from a widely-used rating scale (Barkley & Murphy, 1998), relatively little was known about the psychometric characteristics of these impairment measures prior to the current study. The significant stability of all four impairment measures over a ten-year period provides important evidence

that parent ratings of impairment are reliable and stable over time, and the relation between ADHD symptoms and these impairment measures provides critical support for the concurrent and predictive validity of IN and HI.

While these results support the utility of the current measures of functional impairment, it is also important to acknowledge the limitations of these ratings. Whereas the seven-item scale allowed us to create a latent measure of overall impairment, the social, academic, and recreational impairment measures were composites of measured items that were not free of measurement error or item-specific variance. Further, while stability correlations were significant for all impairment measures over the ten-year period of the study, the stability of several of the functional impairment measures was lower than the stability of the ratings of IN and HI. The lower stability of the functional impairment measures may simply reflect the weaker psychometric properties of these scales, or could potentially indicate that levels of impairment may be more malleable over time than symptoms of ADHD.

Future studies of ADHD and functional impairment should employ measures that assess specific aspects of functional impairment with a larger pool of items or alternative approaches to measure functional impairment. More broadly, systematic research is needed to develop and validate psychometrically sound measures of different dimensions of functional impairment, ideally with adequate normative data to facilitate their use in clinical practice.

Specific limitations of SEM—An important limitation of any structural equation modeling approach is the existence of numerous, equally well-fitting models. The current model provides an acceptable, parsimonious description of the data, and is consistent with the extensive literature that supports a model with two correlated IN and HI factors to describe the structure of ADHD symptoms (Willcutt et al., 2012). However, bi-factor models of ADHD (e.g., Toplak et al., 2012) have also grown in popularity and been shown to demonstrate comparable or superior fit to two-factor models (Willoughby, Blanton, & Family Life Project Investigators, 2015), although recent work questions the validity and interpretability of these models (Eid, Geiser, Koch, & Heene, 2017). These models, in which a general factor is created from all 18 symptoms and IN- and HI-specific factors are created from residual variance in their respective symptoms, have demonstrated, among other things, that although the IN and HI factors are dissociable, their respective symptoms have more shared than unique variance. This emerging literature emphasizes that two-factor, bi-factor, and alternative models are both possible and should be further explored. Finally, it is important to note that these models do not intend to make causal claims about the relationships between ADHD dimensions and impairment. Instead, causality is only implicated by the fact that these reliable predictive effects emerged over time.

Conclusion

Results from a population-based longitudinal sample of 489 twins indicated that the parent-rated IN, HI, and overall impairment constructs were invariant between females and males, as well as across the ten-year period from preschool through the end of ninth grade. Despite small mean differences between males and females, as well as decreasing HI levels over

development, the IN and HI dimensions were highly stability. Latent cross-lagged models indicated that both IN and HI were independently associated with increased risk for concurrent and future overall, social, and recreational impairment, whereas only IN was uniquely associated with later academic impairment. These results underscore the importance of the distinction between symptoms of IN and HI, and suggest that early elevations of IN and HI symptoms may provide key information regarding risk for impairment later in development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4. Washington, DC: American Psychiatric Association; 1994.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5. Washington, DC: American Psychiatric Association; 2013.

Arnett AB, Pennington BF, Willcutt EG, DeFries JC, Olson RK. 2015; Sex differences in ADHD symptom severity. Journal of Child Psychology and Psychiatry. 56:632–639. [PubMed: 25283790]

Barkley, RA, Murphy, K. Attention-deficit hyperactivity disorder: A clinical workbook. 2. New York, NY: Guilford Press; 1998.

Bauermeister JJ, Canino G, Polanczyk G, Rohde LA. 2010; ADHD across cultures: Is there evidence for a bidimensional organization of symptoms? Journal of Clinical Child & Adolescent Psychology. 39(3):362–372. [PubMed: 20419577]

Brosco JP, Bona A. 2017; Changes in academic demands and attention-deficit/hyperactivity disorder in young children. JAMA Pediatrics. 170(4):396–397.

Burns GL, Becker SP, Servera M, Bernad MDM, Garcia-Banda G. 2016Sluggish cognitive tempo and attention-deficit/hyperactivity disorder (ADHD) inattention in the home and school contexts: Parent and teacher invariance and cross-setting validity. Psychological Assessment.

Burns GL, Servera M, Bernad MDM, Carrillo JM, Geiser C. 2014; Ratings of ADHD symptoms and academic impairment by mothers, fathers, teachers, and aides: Construct validity within and across settings as well as occasions. Psychological Assessment. 26(4):1247–1258. [PubMed: 24932644]

Burns GL, Walsh JA, 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. [PubMed: 17154767]

Byrne B, Delaland C, Fielding-Barnsley R, Quain P, Samuelsson S, Hoien T, … Olson RK. 2002; Longitudinal twin study of early reading development in three countries: Preliminary results. Annals of Dyslexia. 52:49–73.

Chen FF. 2007; Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling. 14:464–504.

Chen FF, Sousa KH, West SG. 2005; Testing measurement invariance of second-order factor models. Structural Equation Modeling. 12:471–492.

Christopher ME, Hulsander J, Keenan J, DeFries J, Pennington BF, Byrne B, … Wadsworth S. 2015; Genetic and environmental etiologies of the longitudinal relations between pre-reading skills and reading. Child Development. 86:342–361. [PubMed: 25263167]

Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington BF, … Olson RK. 2013; The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. Journal of Experimental Child Psychology. 125:453–467.

Dittman CK. 2016; The impact of early classroom inattention on phonological processing and word-reading development. Journal of Attention Disorders. 20(8):653–664. [PubMed: 23475828]

Eid M, Geiser C, Koch T, Heene M. 2017; Anomalous results in *g*-factor models: Explanations and alternatives. Psychological Methods. 22(3):541–562. [PubMed: 27732052]

Friedman-Weieneth JL, Doctoroff GL, Harvey EA, Goldstein LH. 2009; The Disruptive Behavior Rating Scale-Parent Version (DBRS-PV): Factor analytic structure and validity among young preschool children. Journal of Attention Disorders. 13:42–55. [PubMed: 18753403]

Gaub M, Carlson CL. 1997; Gender differences in ADHD: a meta-analysis and critical review. Journal of the American Academy of Child and Adolescent Psychiatry. 36:1036–1045. [PubMed: 9256583]

Gershon J. 2002; A meta-analytic review of gender differences in ADHD. Journal of Attention Disorders. 5:143–154. [PubMed: 11911007]

Hartung CM, Lefler EK, Canu WH, Stevens AE, Jaconis M, LaCount PA, … Willcutt EG. 2016*DSM-5* and other symptom thresholds for ADHD: Which is the best predictor of impairment in college students? Journal of Attention Disorders.

Hinshaw SP. 2002; Preadolescent girls with attention-deficit/hyperactivity disorder: I. Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. Journal of Consulting and Clinical Psychology. 70:1086–1098. [PubMed: 12362959]

Hinshaw SP, Owens EB, Sami N, Fargeon S. 2006; Prospective follow-up of girls with attention-deficit/hyperactivity disorder into adolescence: Evidence for continuing cross-domain impairment. Journal of Consulting and Clinical Psychology. 74:489–499. [PubMed: 16822106]

Lahey BB, Pelham WE, Loney J, Kipp H, Ehrhardt A, Lee SS, … Massetti G. 2004; Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in children diagnosed at 4–6 years of age. American Journal of Psychiatry. 161:2014–2020. [PubMed: 15514401]

Lahey BB, Pelham WE, Loney J, Lee SS, Willcutt E. 2005; Instability of the DSM-IV Subtypes of ADHD from preschool through elementary school. Archives of General Psychiatry. 62:896–902. [PubMed: 16061767]

Lahey BB, Pelham WE, Stein MA, Loney J, Trapani C, Nugent K, … Baumann B. 1998; Validity of DSM-IV attention-deficit/hyperactivity disorder for younger children. Journal of the American Academy of Child and Adolescent Psychiatry. 37:695–702. [PubMed: 9666624]

Lahey BB, Willcutt EG. 2010; Predictive validity of a continuous alternative to nominal subtypes of attention-deficit hyperactivity disorder in DSM-IV. Journal of Clinical Child and Adolescent Psychology. 39:761–775. [PubMed: 21058124]

Little, TD. Longitudinal structural equation modeling. New York, NY: Guilford Press; 2013.

Little TD, Rhemtulla M, Gibson K, Schoemann AM. 2013; Why the items versus parcels controversy needn't be one. Psychological Methods. 18:285–300. [PubMed: 23834418]

McGee R, Prior M, Williams S, Smart D, Sanson A. 2002; The long-term significance of teacher-rated hyperactivity and reading ability in childhood: Findings from two longitudinal studies. Journal of Child Psychology and Psychiatry. 43:1004–1017. [PubMed: 12455922]

Muthen, LK, Muthen, BO. Mplus User's Guide. 7. Los Angeles, CA: Muthen and Muthen; 2012.

Olson RK, Keenan JM, Byrne B, Samuelsson S, Coventry WL, Corley R, … Hulslander J. 2011; Genetic and environmental influences on vocabulary and reading development. Scientific Studies of Reading. 15:26–46. [PubMed: 21132077]

Owens EB, Hinshaw SP, Lee SS, Lahey BB. 2009; Few girls with childhood attention-deficit/hyperactivity disorder show positive adjustment during adolescence. Journal of Clinical Child and Adolescent Psychology. 38:132–143. [PubMed: 19130363]

Pham AV. 2016; Differentiating behavioral ratings of inattention, impulsivity, and hyperactivity in children: Effects on reading achievement. Journal of Attention Disorders. 20(8):674–683. [PubMed: 23400214]

Reynolds, CR, Kamphaus, RW. Behavior Assessment System for Children. 2. Circle Pines, MN: American Guidance Service; 2004.

Ros R, Graziano PA. 2017Social functioning in children with or at risk for attention deficit/ hyperactivity disorder: A meta-analytic review. Journal of Clinical Child and Adolescent Psychology.

Spitzer RL, Wakefield JC. 1999; DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? American Journal of Psychiatry. 156:1856–1864. [PubMed: 10588397]

Toplak ME, Sorge GB, Flora DB, Chen W, Banaschewski T, Buitelaar J, … Faraone SV. 2012; The hierarchical factor model of ADHD: Invariant across age and national groupings? Journal of Child Psychology and Psychiatry. 53(3):292–303. [PubMed: 22084976]

Willcutt EG. 2012; The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics. 9:490–499. [PubMed: 22976615]

Willcutt EG, Betjemann RS, McGrath LM, Chhabildas NA, Olson RK, DeFries JC, Pennington BF. 2010; Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. Cortex. 46:1345–1361. [PubMed: 20828676]

Willcutt EG, Betjemann RS, Pennington BF, Olson RK, DeFries JC, Wadsworth SJ. 2007; Longitudinal study of reading disability and attention-deficit/hyperactivity disorder: Implications for education. Mind, Brain, and Education. 1(4):181–192. DOI: 10.1111/j.1751-228X. 2007.00019.x

Willcutt EG, Betjemann RS, Wadsworth SJ, Samuelsson S, Corley R, DeFries JC, … Olson RK. 2007; Preschool twin study of the relation between attention-deficit/hyperactivity disorder and prereading skills. Reading and Writing. 20:103–125.

Willcutt EG, Boada R, Riddle MW, Chhabildas N, DeFries JC, Pennington BF. 2011; Colorado Learning Difficulties Questionnaire: Validation of a parent-report screening measure. Psychological Assessment. 23:778–791. [PubMed: 21574721]

Willcutt EG, Nigg JT, Pennington BF, Solanto MV, Rohde LA, Tannock R, … Lahey BB. 2012; Validity of DSM-IV attention-deficit/hyperactivity disorder dimensions and subtypes. Journal of Abnormal Psychology. 121:991–1010. [PubMed: 22612200]

Willoughby MT, Blanton ZE. Family Life Project Investigators. 2015; Replication and external validation of a bi-factor parameterization of attention deficit/hyperactivity symptomatology. Journal of Clinical Child & Adolescent Psychology. 44:68–79. [PubMed: 24256437]
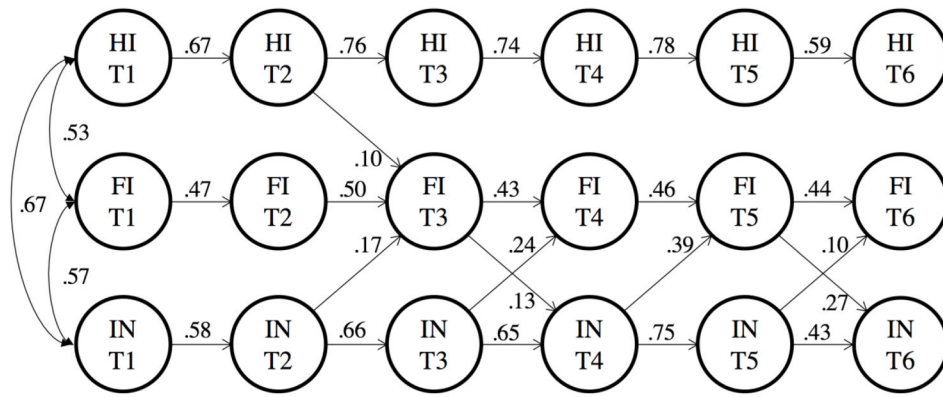
Fig. 1.
Structural equation model of the latent cross-lagged relationships between inattention (IN), hyperactivity/impulsivity (HI), and overall functional impairment. $N = 489$. All cross-lagged paths of IN and HI predicting the impairment domain, and the impairment domain predicting IN and HI were included, but only significant ($p < .05$) paths are depicted. Lines with single arrowheads represent directed regression paths, whereas lines with double arrowheads represent bivariate correlations. The disturbance terms at each time point after preschool are significantly correlated with one another (see Supplemental Table 3). IN = inattention; HI = hyperactivity/impulsivity; FI = functional impairment; Imp. = impairment; T1 – T6 = preschool, post K, post 1st, post 2nd, post 4th, and post 9th grade.

Table 1

Descriptive Statistics, Internal Reliability, and Longitudinal Correlations of Manifest IN, HI, and Overall Impairment Parent Ratings

Measure/ Wave	M(SD)	Female M(SD)	Male M(SD)	α	Longitudinal correlations within dimensions				
					Post K	Post 1st	Post 2nd	Post 4th	Post 9th
Inattention									
Preschool	.67 (.50)	.61 (.46)	.73 (.53)	.89	.57	.50	.46	.46	.47
Post K	.56 (.48)	.47 (.44)	.65 (.50)	.89	--	.65	.63	.56	.44
Post 1st	.57 (.50)	.50 (.44)	.64 (.54)	.90			.67	.64	.47
Post 2nd	.59 (.51)	.50 (.44)	.68 (.56)	.90				.72	.52
Post 4th	.65 (.56)	.56 (.53)	.74 (.58)	.92					.60
Post 9th	.60 (.60)	.49 (.56)	.70 (.61)	.93					
Hyperactivity-Impulsivity									
Preschool	.78 (.55)	.68 (.49)	.88 (.60)	.87	.64	.56	.58	.51	.44
Post K	.55 (.51)	.45 (.43)	.65 (.55)	.88		.67	.70	.64	.39
Post 1st	.49 (.49)	.41 (.39)	.57 (.56)	.88			.66	.66	.40
Post 2nd	.45 (.46)	.36 (.37)	.55 (.53)	.88				.71	.53
Post 4th	.40 (.44)	.33 (.37)	.48 (.50)	.87					.50
Post 9th	.31 (.41)	.26 (.33)	.36 (.47)	.86					
Overall Impairment									
Preschool	.52 (.47)	.47 (.48)	.58 (.47)	.86	.51	.47	.40	.46	.39
Post K	.58 (.55)	.50 (.51)	.65 (.58)	.89		.57	.63	.56	.46
Post 1st	.55 (.53)	.48 (.49)	.60 (.56)	.88			.52	.53	.37
Post 2nd	.59 (.55)	.48 (.47)	.70 (.61)	.90				.68	.39
Post 4th	.55 (.51)	.50 (.49)	.61 (.53)	.87					.50
Post 9th	.64 (.60)	.56 (.57)	.71 (.61)	.89					--

Note. Means and standard deviations on 0 – 3 Likert scale. IN = inattention; HI = hyperactivity/ impulsivity. All sex differences significant, $p < .05$. All correlations are significant, $p < .01$.

Author Manuscript

Table 2

Manifest Correlations Between IN, HI, and Overall Impairment at Each Assessment

	Assessment					
	Preschool	Post K	Post 1st	Post 2nd	Post 4th	Post 9th
Correlations with IN						
Hyperactivity-Impulsivity	.59	.55	.57	.57	.56	.52
Overall impairment	.50	.55	.62	.61	.70	.66
Social Impairment	.41	.44	.49	.51	.55	.49
Academic Impairment	.48	.58	.64	.61	.69	.70
Recreational Impairment	.31	.36	.39	.42	.49	.46
Correlations with HI						
Overall impairment	.47	.52	.53	.52	.54	.47
Social impairment	.45	.48	.48	.52	.52	.46
Academic impairment	.32	.40	.40	.38	.36	.35
Recreational impairment	.32	.44	.40	.44	.43	.36

Note. Correlations reflect the manifest associations between mean levels of IN, HI, and impairment at each assessment. IN = inattention; HI = hyperactivity/impulsivity; all p <.001.

Table 3

Invariance of Parent-rated IN, HI, and FI Symptoms at Preschool, Post K, Post 1st, Post 2nd, Post 4th, and Post 9th Grades

Model tested	df	χ²	p	df	SBχ²	p	CFI	TLI	RMSEA [90% CI]	MC	CFI	TLI	Pass?
Temporal invariance models using item-level indicators													
1: Configural invariance	4634	6624.0	<.001	–	–	–	.94	.94	.030 [.028, .031]	–	–	–	Yes
2: Strong invariance	4838	6868.6	<.001	–	–	–	.94	.94	.029 [.028, .031]	2 vs. 1	–.001	+.002	Yes
Multiple group sex invariance models using item-level indicators													
1: Configural invariance	9458	11207.0	<.001	–	–	–	.94	.94	.028 [.025, .030]	–	–	–	Yes
2: Strong invariance	9788	11395.2	<.001	–	–	–	.95	.95	.026 [.024, .028]	2 vs. 1	+.002	+.006	Yes
Multiple group sex invariance models using parcelled indicators													
1: Configural invariance	2178	2907.6	<.001	–	–	–	.96	.95	.037 [.033, .041]	–	–	–	Yes
2: Weak invariance	2244	2966.3	<.001	–	–	–	.96	.95	.036 [.033, .040]	2 vs. 1	+.001	+.002	Yes
3: Strong invariance	2310	3075.6	<.001	–	–	–	.96	.95	.037 [.033, .040]	3 vs. 1	–.001	.000	Yes
SEM comparisons across sex													
4: Means - omnibus	2328	3114.8	<.001	18	39.8	.002	.96	.95	.037 [.034, .041]	4 vs. 3	–	–	No
5: Means - IN & FI (HI free to vary)	2322	3102.9	<.001	12	27.7	.006	.96	.95	.037 [.034, .040]	5 vs. 3	–	–	Yes/No
6: Means - HI & FI (IN free to vary)	2322	3106.4	<.001	12	32.0	.001	.96	.95	.037 [.034, .041]	6 vs. 3	–	–	No
7: Means - IN & HI (FI free to vary)	2322	3106.5	<.001	12	31.2	.002	.96	.95	.037 [.034, .041]	7 vs. 3	–	–	No
8: Variance constraints	2328	3108.8	<.001	18	30.0	.038	.96	.95	.037 [.034, .040]	8 vs. 3	–	–	Yes
9: Variance/covariance constraints	2346	3141.5	<.001	36	59.8	.008	.96	.95	.037 [.034, .041]	9 vs. 3	–	–	Yes

Note. IN = inattention; HI = hyperactivity/impulsivity; FI = functional impairment; SBχ2 = Satorra-Bentler chi-square; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; CI = confidence interval; MC = model comparison; Pass = comparison meets invariance study criteria.

Table 4

Descriptive Statistics, Sex Differences, and Longitudinal Change of Latent IN, HI, and Overall Impairment Parent Ratings

Measure/ Wave	M(SD)	Female M(SD)	Male M(SD)	d	Longitudinal change (d) within dimensions				
					Post K	Post 1st	Post 2nd	Post 4th	Post 9th
Inattention									
Preschool	.66 (.47)	.61 (.44)	.72 (.50)	.25	-.22	-.18	-.13	-.02	-.13
Post K	.56 (.45)	.47 (.42)	.65 (.47)	.40	--	.04	.08	.18	.07
Post 1st	.58 (.49)	.51 (.44)	.65 (.52)	.30		--	.04	.14	.03
Post 2nd	.60 (.49)	.53 (.45)	.68 (.53)	.32			--	.10	-.01
Post 4th	.65 (.54)	.56 (.52)	.75 (.55)	.34				--	-.10
Post 9th	.60 (.58)	.51 (.56)	.69 (.59)	.32					--
Hyperactivity-Impulsivity									
Preschool	.78 (.52)	.68 (.45)	.88 (.56)	.40	-.46	-.57	-.68	-.78	-1.06
Post K	.55 (.47)	.46 (.41)	.66 (.51)	.43	--	-.11	-.21	-.32	-.59
Post 1st	.50 (.47)	.42 (.39)	.58 (.53)	.34		--	-.10	-.20	-.48
Post 2nd	.46 (.43)	.37 (.35)	.55 (.49)	.43			--	-.11	-.39
Post 4th	.41 (.43)	.34 (.35)	.48 (.48)	.35				--	-.28
Post 9th	.30 (.38)	.25 (.31)	.35 (.44)	.24					--
Overall Impairment									
Preschool	.50 (.43)	.46 (.44)	.54 (.42)	.17	.07	.04	.11	.03	.16
Post K	.53 (.51)	.46 (.48)	.59 (.53)	.25	--	-.03	.04	-.04	.08
Post 1st	.52 (.50)	.47 (.47)	.58 (.53)	.21		--	.07	-.01	.11
Post 2nd	.56 (.53)	.46 (.46)	.65 (.57)	.37			--	-.09	.04
Post 4th	.51 (.48)	.46 (.45)	.55 (.50)	.19				--	.13
Post 9th	.58 (.56)	.51 (.54)	.63 (.78)	.21					--

Note. Means and standard deviations on 0 – 3 Likert scale. Latent Cohen's *d* calculated according to Little (2013) p. 170. IN = inattention; HI = hyperactivity/impulsivity. Sex differences significant, *p* <.05.