CrossMark

Examination of the Structure and Measurement of Inattentive, Hyperactive, and Impulsive Behaviors from Preschool to Grade 4

Darcey M. Allan[1] · Christopher J. Lonigan[2]

Abstract

Attention Deficit/Hyperactivity Disorder (ADHD) is a complex and heterogeneous disorder consisting of inattentive and hyperactive/impulsive behaviors. Although, the multidimensionality of ADHD is widely accepted, questions remain regarding the extent to which the components of this disorder are overlapping or distinct. Further, although the same measures are generally used to assess inattentive, hyperactive, and impulsive behaviors across childhood, it has been argued that the structure and measurement of inattentive, hyperactive, and impulsive behaviors may be susceptible to developmental influences. The purpose of this study was to examine the factor structure and measurement invariance of inattentive and hyperactive/impulsive behaviors in a large group of children ($N = 10,047$) ranging in grade level from preschool to grade 4. A bifactor model with a general factor and two specific factors of inattention and hyperactivity/impulsivity fit the data best. This finding held across all groups and all grade levels. In general, the bifactor model demonstrated measurement invariance from kindergarten through grade 4 but not for preschool. Implications for the understanding and measurement of inattentive, hyperactive, and impulsive behaviors across early and middle childhood are discussed.

Keywords Inattention · Hyperactivity/impulsivity · Bifactor model · Factor structure · Invariance · Developmental

Attention Deficit/Hyperactivity Disorder (ADHD) is a complex and heterogeneous developmental disorder consisting of inattentive and hyperactive/impulsive behaviors that affects approximately 3 to 7% of school-age children (American Psychiatric Association [APA] 2013). Although, the multidimensionality of ADHD is widely accepted, questions remain regarding the extent to which the components of this disorder are overlapping or distinct. Further, although the general behavioral descriptions that make up the diagnostic criteria for ADHD are the same across childhood, it has been argued that

the structure and measurement of inattentive, hyperactive, and impulsive behaviors may be susceptible to developmental influences (Lahey et al. 2005). Thus, studies that address issues of measurement using samples of children at a range of developmental levels are needed. The purpose of this study was to determine, in a large sample of children, if there were developmental differences in the structures and relations of the inattentive, hyperactive, and impulsive behaviors of ADHD from preschool to grade 4.

✉ Darcey M. Allan
dallan@ohio.edu

✉ Christopher J. Lonigan
lonigan@psy.fsu.edu

[1] Department of Psychology, Ohio University, 200 Porter Hall, Athens, OH 45701, USA

[2] Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee, FL 32306-4301, USA

Structure of ADHD

Although the concept of the disorder currently referred to as ADHD has existed for more than a century, the underlying structure of the behaviors associated with ADHD is still not completely clear. ADHD has been conceptualized differently across versions of the Diagnostic and Statistical Manual (DSM; e.g., APA 1994, 2013). Although ADHD was initially viewed as a unidimensional disorder (i.e., Hyperkinetic Reaction of Childhood; APA 1968), later versions of the DSM have defined the disorder as a multi-dimensional construct. The current conceptualization of ADHD encompasses

three distinct but closely related problem behavior areas: inattention, hyperactivity, and impulsivity. Behaviors within these areas are listed as separate symptoms; however, hyperactivity and impulsivity traditionally have been combined to reflect a single diagnostic construct.

Clarifying the structure of these closely related problem behaviors is important to the evolution of how ADHD and its various presentations are conceptualized and classified. The structural conceptualizations of these disorders are relevant to the diagnosis, etiology, and treatment of the problem behaviors associated with ADHD. Findings related to the structure of these symptoms may guide the effective use of subtypes and specifiers or assist in the identification of ADHD-related symptoms that may represent distinct disorders (e.g., sluggish cognitive tempo). Furthermore, research examining the unique and shared variance among ADHD symptoms informs an understanding of which symptoms do and do not share the same underlying causes. Perhaps most importantly, research related to the potential distinctness of certain ADHD symptom clusters may lead to different treatment approaches for children who exhibit different symptom presentations.

Across the four most recent editions of the DSM, hyperactivity and impulsivity have been characterized as a single behavioral dimension. A number of factor analytic studies of ADHD based on both parent and teacher ratings have supported the presence of two factors: Inattention and Hyperactivity/Impulsivity (H/I; e.g., Molina et al. 2001; DuPaul et al. 1997; Wolraich et al. 2003). Although some evidence supports potential utility in parsing hyperactivity and impulsivity, researchers suggest that the two constructs should remain together based on the rule of parsimony. For example, Toplak et al. (2009) examined the factor structure of ADHD based on categorical item-level data and found evidence of similar fit between a two-factor and a three-factor models; they determined that a bifactor model with two specific factors (i.e., inattention and H/I) to be the best-fitting model based on parsimony. Similarly, in an analysis of the factor structure of both ADHD and Oppositional Defiant Disorder (ODD), Burns et al. (2001) reported that a 4-factor model (ODD, Inattention, Hyperactivity, and Impulsivity) in which hyperactivity and impulsivity were modeled as distinct factors fit the data best. However, because the incremental improvement in fit was minimal and the correlations between the hyperactivity and impulsivity factors were high, they determined the more parsimonious model with hyperactivity and impulsivity items modeled on the same factor to be the preferred model.

Given the substantial associations among the distinct dimensions of ADHD, research has also examined the hierarchical structure of ADHD. Ullebø et al. (2012) used bifactor models to examine the dimensionality of ADHD in a community sample of children in grades 2 to 4 using both parent and teacher reports of ADHD. They tested both second-order CFA and bifactor models and reported that a bifactor model with a general ADHD factor and specific factors for inattention and impulsivity provided the best fit for the data. In their model, there was not support for a specific hyperactivity factor. Martel et al. (2011) reported on a bifactor model of ADHD in a group of children and adolescent boys, and they modeled hyperactivity and impulsivity as a single specific factor. Wagner et al. (2015) examined the structure of ADHD in a sample of high-risk children and adolescents using both correlated-traits and bifactor models. Although they reported that a bifactor model with a General ADHD factor and three specific factors (i.e., inattention, hyperactivity, and impulsivity), analyses concerning the reliability of the distinct factors indicated that the specific factors did not account for substantial variance above and beyond that which was accounted for by the General ADHD factor. In a sample of younger children (4 to 6 years of age) using bifactor exploratory structural equation modeling, Arias et al. (2016) found support for a model with distinct inattention, hyperactivity, and impulsivity factors. Again, the weak reliability of the specific factors challenged the meaningfulness of the specific factors.

Taken together, these findings highlight some of the mixed evidence concerning the separation of hyperactivity and impulsivity as they related to ADHD. The use of bifactor modeling has introduced an intriguing way of characterizing behaviors, such as those associated with ADHD that comprises both overlapping and unique components. Although extant bifactor analyses have consistently identified a General ADHD factor and a Specific Inattention Factor (e.g., Arias et al. 2016; Gómez et al. 2017; Wagner et al. 2015; Willoughby et al. 2015), questions remain regarding whether to model hyperactivity and impulsivity as distinct or combined behavioral constructs.

A Developmental Approach to ADHD

Another issue related to the conceptualization of attention, hyperactivity, and impulsivity is the underdeveloped understanding of how these behaviors and the relations between these behaviors change across development. With some exceptions, the DSM generally operates from a descriptive symptom-based approach that bases disorders on the presence of clearly observable or reportable behaviors (APA 2013). The selection and wording of symptom criteria is particularly critical for disorders, such as ADHD, that appear in early childhood and often persist into adulthood (Barbaresi et al. 2013). The symptoms that individuals experience and the manifestations of these symptoms may vary over the course of development. Some behaviors may be more or less developmentally appropriate at different ages. Some descriptions of inattention and H/I are based on English language idioms (e.g., driven by a motor) that may evoke different interpretations depending on the age of the child. Moreover, the overt

manifestations of inattentive, hyperactive, and impulsive behaviors may change over the course of development (Biederman et al. 2000; Larsson et al. 2011).

Using the DSM IV-TR and DSM-5, the same general symptoms are used to determine the presence or absence of ADHD in all individuals regardless of age. In the DSM-5, efforts have been made to facilitate consideration of an individual's developmental level in the process of symptom identification. For example, when considering the presence, absence, severity, and frequency of symptoms on rating scales, informants are typically asked to consider whether the behavior is typical of children who are the same age. To help account for changing manifestations of symptoms across development, the DSM-5 includes examples regarding symptom presentation at different age groups. However, most of these considerations address manifestations of ADHD in adulthood. For example, the symptom "often runs about…" in children is noted to be possibly limited to "restlessness" in adults. Distinctions are typically not made between the manifestations of behaviors in early childhood and middle childhood; however, these distinctions may be important. For example, excessive motor activity may present as rolling on the floor during circle time for preschoolers, but it may present as foot-tapping and hand-fidgeting for fourth-graders.

Given that ADHD is a multidimensional disorder, developmental differences may also exist at the construct level. There is substantial longstanding evidence that hyperactivity emerges early in childhood and often subsides over the course of development, whereas inattention may not be easily discernable in preschool, but may become the more prominent symptom cluster once children reach middle childhood and adolescence (Biederman et al. 2000; Larsson et al. 2011). Moreover, specific behaviors presumed to reflect one of these symptom clusters may not be equally indicative of psychopathology across different developmental stages. Importantly, considering the age of the child and using developmentally-appropriate examples of behavior in symptom descriptions does not necessarily change the possibility that certain behaviors may be more or less indicative of significant problems with inattention, hyperactivity, or impulsivity at different points in development. For example, the DSM-5 impulsivity symptom "often interrupts or intrudes on others," as listed in the DSM, offers several examples of how this behavior may manifest at different developmental levels. However, impulsively taking a toy from another child may not be a behavior of particular concern for preschoolers who are just learning to share and take turns, but interrupting activities being completed by another child may be a stronger indicator of symptomology for third-graders for which sharing and turn-taking are well-established social norms. Thus, although these behaviors reflect the same symptom at different developmental levels, they may be differentially related to overall levels of ADHD-related problem behaviors.

Concerns related to the developmental changes in ADHD presentation (Coghill and Seth 2011; Frick and Nigg 2012) have led to a growing body of research examining the measurement of inattention and H/I across a wide developmental range. The assessment of inattentive, hyperactive, and impulsive behaviors during the preschool period may be vulnerable to lack of measurement invariance, particularly when the ratings of behavior are made by teachers who observe the child in the context of the preschool classroom. Preschool may be a child's first exposure to structured group activities during which the self-regulation of attention, motor activity, and impulses is critical. Moreover, compared to elementary-level teachers, preschool teachers vary in terms of their training, experience, and expectations for student behavior. These variations in child behavior and teacher factors may impact ratings of student behaviors.

An increasing number of studies have examined the factor structure of ADHD in samples that include young children. Many studies have supported measurement invariance from early childhood through adolescence (e.g., Caci et al. 2016; Leopold et al. 2016). However, extant studies examining the factor structure of inattention and H/I across development either do not include preschool children (e.g., DuPaul et al. 2016), do not test age-related measurement invariance at the preschool level specifically (e.g., Caci et al. 2016), or do not examine measurement invariance using bifactor models (e.g., Leopold et al. 2016; McGoey et al. 2015).

Summary and Purpose

Despite substantial advancements in developmentally-sensitive conceptualizations of ADHD, questions remain regarding the structure and measurement of inattentive, hyperactive, and impulsive behaviors across different developmental periods. Bifactor modeling offers a useful way of examining both the unique and overlapping aspects of the behaviors associated with ADHD. The primary purpose of this study was to examine the structure and measurement of inattentive and hyperactive/impulsive behaviors in a large community-based sample of children in early and middle childhood. The first goal of the study was to examine the factor structure of inattention, hyperactivity, and impulsivity across this age span using confirmatory factor analysis (e.g., correlated-traits models and bifactor models). Given that ADHD is currently conceptualized as a unitary disorder with different presentations (American Psychiatric Association 2013), it was expected that a bifactor model comprised of a General ADHD Factor and two specific factors (i.e., Inattention and H/I) would provide the best fit to the data. The second goal of the study was to examine whether the measurement of inattention and H/I was the same from preschool through grade 4. Given that the same general behaviors are used to assess ADHD across childhood (American Psychiatric Association 2013), it was

expected that the measurement of inattention, hyperactivity, and impulsivity would be the same across grades.

Method

Participants

As a part of three larger studies, children were recruited from schools that served primarily children from low-income families in north Florida. The sample included a total of 10,047 children (Preschool $n = 1532$; Kindergarten $n = 2404$; Grade 1 $n = 1916$; Grade 2 $n = 1618$; Grade 3 $n = 1248$; Grade 4 $n = 1618$). The sample was 50.3% female and was racially diverse (i.e., 59% White, 29% African-American/Black, 12% other ethnicities). Results of a χ^2 test, $\chi^2 = 7.71$ (5), $p = .17$, indicated that the distributions of gender were similar across grades. However, a χ^2 test, $\chi^2 = 30.84$ (5), $p < 0.001$, indicated that there was a higher proportion of students from racial minority backgrounds in the younger grades compared to the older grades (Preschool = 43.1%, Kindergarten = 43.6%, Grade 1 = 41.2%, Grade 2 = 40.6%, Grade 3 = 35.9%, Grade 4 = 37.2%). Across grades, mean age was 56.02 months ($sd = 4.61$) in preschool, 68.34 months ($sd = 4.94$) in kindergarten, 80.90 months ($sd = 5.74$) in Grade 1, 93.21 months ($sd = 6.38$) in Grade 2, 106.17 months ($sd = 6.95$) in Grade 3, and 118.79 months ($sd = 7.01$) in Grade 4.[1] Although specific data related to socioeconomic status was not available, all schools from which participants were recruited served a high proportion of children who are eligible for free and reduced price lunch.

Measures

The Strengths and Weaknesses of ADHD-Symptoms and Normal-Behaviors Rating Scale (SWAN) The SWAN (Swanson et al. 2001) includes 27 items that correspond to the diagnostic criteria for ADHD and Oppositional Defiant Disorder as described in the Diagnostic Statistical Manual Text Revision 4th Edition (American Psychiatric Association 2000). Children were rated based on comparisons to same-age peers, and scores ranged from −3 to 3 for each item. The SWAN has been shown to have strong internal consistency (alpha values are greater than or equal to .95) and test-retest reliability (rs between .71 and .76) for each subscale (Lakes et al. 2012). The measure has also been shown to have good validity as

evidenced by strong correlations with other established measures of ADHD (e.g., Arnett et al. 2013). In this sample, the SWAN also showed a high level of internal consistency ($\alpha = 0.98$), with good reliability for the inattention items ($\alpha = 0.98$), the hyperactivity items ($\alpha = 0.96$), and the impulsivity items ($\alpha = 0.93$). Reliability and validity has been demonstrated in children as young as 3-years-old (Lakes et al. 2012).

Procedure

Data for this study came from three larger studies (e.g., see Connor et al. 2018; Lonigan and Burgess 2017; Lonigan and Milburn 2017; Lonigan et al. 2018) investigating the development of reading-related skills or evaluating instructional approaches to improve reading-related skills. Procedures for all projects were approved by Florida State University's Institutional Review Board. At participating schools, parents were invited to participate via parental consent forms sent home by their child's teacher. Across all studies, informed consent was obtained from children's parents or legal guardians before data collection began. The children's primary teachers were asked to complete the SWAN within the first two to three months of the school year, and teachers completed their ratings prior to any additional instruction conducted as part of the intervention project for children in the intervention projects. Teachers completed the SWAN either as a pen-and-paper questionnaire or as an online questionnaire. Teacher ratings were completed by 1155 teachers working in 113 schools within 11 school districts in North Florida. Across projects, teachers received nominal monetary compensation for the time required to complete the SWAN.

Results

Descriptive statistics for all items on the SWAN are presented in Table 1. Across items, data was missing for less than 0.1% of the data. Overall, the means reflect those of an average sample. Because of the large sample available in each grade and to prevent identification of statistically significant but trivial differences, children within each grade were divided into three randomly derived subsamples (as determined based on random number generation in Microsoft Excel). This approach allowed for cross-subsample replication of each test of factor structure and each test of measurement invariance. Because of the large sample size and the number of multiple comparisons, Bonferroni corrections were used for each set of comparisons within subsample to provide more conservative estimates of statistical significance.

Confirmatory Factor Analyses

CFAs were conducted using Mplus version 7.3 (Muthén and Muthén 2014). Full information maximum likelihood was used

---

[1] The potential impact of within-grade age on the findings was examined by calculating the amount of variance in item-level scores accounted for by age within grade. With few exceptions, age accounted for less than 1% in item-level scores across items. For most items, the amount of variance accounted for by age was less than .04%. Age ranges are as follows: Preschool: 39 months to 103 months; Kindergarten: 52 months to 107 months; Grade 1: 54 months to 107 months; Grade 2: 59 months to 124 months; 66 months to 139 months; Grade 4: 95 months to 148 months.

Table 1 Descriptive statistics for SWAN items across grade

Item	PK[a] M	PK[a] SD	K[b] M	K[b] SD	Grade 1[c] M	Grade 1[c] SD	Grade 2[d] M	Grade 2[d] SD	Grade 3[e] M	Grade 3[e] SD	Grade 4[f] M	Grade 4[f] SD
1	−0.14	1.39	−0.10	1.55	−0.10	1.67	−0.13	1.67	−.17	1.60	−0.30	1.57
2	−0.04	1.32	−0.03	1.50	−0.04	1.52	−0.02	1.61	−.02	1.52	−0.13	1.50
3	0.14	1.28	0.24	1.36	0.27	1.40	0.36	1.47	.43	1.35	0.26	1.34
4	0.03	1.38	0.02	1.51	0.07	1.58	0.14	1.65	.13	1.58	−0.07	1.59
5	−0.09	1.30	−0.03	1.45	−0.02	1.53	0.00	1.60	.01	1.54	−0.12	1.53
6	−0.06	1.35	0.03	1.51	0.07	1.60	0.07	1.67	0.05	1.58	−0.09	1.57
7	−0.01	1.28	0.03	1.45	0.07	1.50	0.09	1.57	0.03	1.52	−0.03	1.53
8	−0.37	1.23	−0.34	1.44	−0.32	1.45	−0.29	1.58	−0.33	1.49	−0.37	1.52
9	0.31	1.27	0.29	1.31	0.36	1.39	0.35	1.45	0.26	1.37	0.13	1.39
10	−0.10	1.37	−0.08	1.49	−0.01	1.53	0.08	1.57	0.10	1.46	0.09	1.43
11	0.04	1.34	0.02	1.48	0.11	1.50	0.23	1.55	0.25	1.43	0.21	1.41
12	0.09	1.23	0.06	1.32	0.17	1.33	0.37	1.44	0.29	1.28	v.33	1.30
13	0.11	1.26	0.06	1.36	0.18	1.39	0.32	1.46	0.30	1.33	0.29	1.33
14	0.13	1.26	0.03	1.37	0.14	1.38	0.26	1.45	0.28	1.32	0.25	1.31
15	−0.01	1.26	−0.14	1.45	−0.06	1.46	0.06	1.54	0.05	1.43	0.05	1.45
16	0.01	1.29	0.04	1.47	0.12	1.49	0.19	1.56	0.19	1.44	0.14	1.47
17	0.11	1.26	0.20	1.36	0.32	1.37	0.38	1.47	0.40	1.35	0.34	1.34
18	−0.03	1.24	0.05	1.33	0.12	1.35	0.22	1.47	0.20	1.34	0.15	1.33

SWAN, The Strengths and Weaknesses of ADHD-Symptoms and Normal-Behaviors Rating Scale

[a] $n = 1532$

[b] $n = 2404$

[c] $n = 1916$

[d] $n = 1618$

[e] $n = 1248$

[f] $n = 1329$

to account for missing data. The Yuan-Bentler scaled chi-square (Y-B χ2) was used to account for the lack of independence between the variables (Yuan and Bentler 2000). Because children were nested within classrooms, robust standard errors were calculated using a sandwich estimator (Muthén and Satorra 1995). Models were first examined for overall model fit using the Y-B χ², the Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC), the comparative fit index (CFI), the Tucker Lewis Index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). The criteria for a good-fitting model included a nonsignificant Y-B χ², a CFI value equal to or greater than .95 and a RMSEA value below .05 (Hu and Bentler 1999; MacCallum et al. 1996). SRMR values close to .08 and below suggest adequate fit (Hu and Bentler 1999; MacCallum et al. 1996). Because achieving a nonsignificant Y-B χ² is a common problem for analyses using large samples, greater weigh was placed on CFI, RMSEA, and SRMR values than on the Y-B χ². Nested models were compared sequentially using chi-square difference testing. A significant chi-square difference indicated that the nested (less parsimonious) model provided improved fit to the data compared to the more parsimonious model. Lower BIC and AIC values also indicated better model fit.

Several a priori hypothesized correlated traits CFA models were compared sequentially. Model comparisons were made separately for each subsample within each grade level. The baseline model consisted of all items loading on a single factor. This model was compared to a two-factor model with an Inattention factor comprising the nine items representing the symptoms of inattention and a H/I factor comprising the six items representing the symptoms of hyperactivity and the three items that represent impulsivity. This two-factor model was compared to a three-factor model in which each of the attention, hyperactivity, and impulsivity constructs was modeled as a separate factor. The differences in chi-square values, all of which were significant, and changes in CFI values comparing the 1-factor correlated-traits model, the 2-factor correlated-traits model, and the 3-factor correlated-traits model are listed by grade and by group in Table 2. A complete report of all fit statistics and comparisons between 1-factor, 2-factor, and 3-factor models by grade and subsample are available in Section A of the online supplemental materials. Across all subsamples and grade-levels, the 3-factor model was determined to be the best-fitting correlated-traits model, and it demonstrated generally good model fit. The correlations between factors in the 3-factor model were large. The correlations between the Inattention factor and the Hyperactivity factor ranged between .68 and .82 across subsample and grade; the correlation between the Inattention factor and the Impulsivity factor ranged between .72 and .82 across subsample and grade; the correlation between the Hyperactivity factor

and the Impulsivity factor ranged between .91 and .96 across subsample and grade.

Given the strong correlations between the Hyperactivity and Impulsivity factors in the 3-factor correlated-traits model, omegas were computed to examine the reliability of the variance explained by each scale in both the 2-factor and 3-factor correlated-traits models; results are presented in Table B1 in the supplemental materials. In the 3-factor correlated-traits model, across grade level and group, both the Inattention and Hyperactivity factors demonstrated excellent reliability (Ωs between .93 and .98); however, the reliability of the Impulsivity factor was below what is considered adequate (Ωs between .38 and .49). In the 2-factor correlated-traits model, across grade level and groups, both the Inattention factor and H/I factor demonstrated adequate reliability (Ωs between 0.80 and 0.98). Taken together, the strong interrelatedness of the Hyperactivity and Impulsivity factors and the poor reliability of the impulsivity items when modeled as a separate factor, the more parsimonious 2-factor correlated-traits model was selected as the best-fitting correlated-traits model.

Following the selection of the best-fitting correlated-traits model, the 2-factor correlated traits model was then compared to a bifactor model. The bifactor model consisted of two orthogonal specific Inattention and H/I factors and a General ADHD factor that modeled the variance shared among all the items. The results for comparisons between the 2-factor correlated-traits model and the bifactor model are shown in Table 2. Across all grade levels and subsamples, the bifactor model fit the data better

than did the two-factor correlated-traits model as evidenced by significant chi-square difference tests and change in CFI equal or greater than .01. The model fit indices for the bifactor model for each grade level and group are shown in Table B2. Across grades and subsamples, with the exception of RMSEA values greater than .05, the bifactor models demonstrated generally good model fit (i.e., CFIs > 0.95, SRMR < 0.08).

Invariance Testing

Because the bifactor model was the best fitting model across grades and subsamples, invariance testing was conducted using this model. Two forms of measurement invariance were tested in the present study: metric invariance and scalar invariance. Metric invariance is used to determine whether a factor assesses the same underlying latent construct (i.e., the latent variable has the same meaning) across groups and is tested by examining whether factor loadings of items are equivalent across groups. Scalar invariance is used to determine whether means of specific item scores are equivalent across groups at the same level of a latent construct (i.e., the latent variable has the same effect on item-level scores across groups) and is tested by examining whether item intercepts are equal across groups. Scalar invariance indicates that means on the latent variable can be compared across groups. In the present study, scalar invariance was only tested if metric invariance was established.

Table 2 Chi-Square differences between models across grade and groups

	ΔChi-Square						ΔCFI					
	Group 1						Group 1					
	PK	K	Grade 1	Grade 2	Grade 3	Grade 4	PK	K	Grade 1	Grade 2	Grade 3	Grade 4
Model 1 vs. 2	178.96	--[a]	1656.73	115.77	147.52	1514.48	0.15	0.20	0.22	0.27	0.24	0.20
Model 2 vs. 3	56.40	103.01	40.35	25.55	17.99	54.12	0.01	0.02	0.01	0.01	0.01	0.01
Model 2 vs. BF	158.87	346.52	243.26	138.94	210.84	218.98	0.03	0.04	0.04	0.02	0.04	0.03
	Group 2						Group 2					
	PK	K	Grade 1	Grade 2	Grade 3	Grade 4	PK	K	Grade 1	Grade 2	Grade 3	Grade 4
Model 1 vs. 2	130.05	2235.04	--[a]	205.79	279.04	105.06	0.16	0.22	0.27	0.28	0.19	0.18
Model 2 vs. 3	24.56	77.01	52.34	83.98	43.70	39.26	0.01	0.01	0.01	0.01	0.01	0.02
Model 2 vs. BF	180.94	310.50	185.67	324.85	226.92	228.64	0.05	0.04	0.02	0.05	0.04	0.05
	Group 3						Group 3					
	PK	K	Grade 1	Grade 2	Grade 3	Grade 4	PK	K	Grade 1	Grade 2	Grade 3	Grade 4
Model 1 vs. 2	248.01	629.32	534.37	322.96	129.23	411.34	0.16	0.20	0.20	0.28	0.27	0.23
Model 2 vs. 3	46.85	77.16	53.31	19.12	31.74	26.64	0.01	0.01	0.01	0.00	0.01	0.00
Model 2 vs. BF	234.41	297.78	226.85	129.30	211.17	198.03	0.05	0.03	0.04	0.02	0.04	0.03

Group 1 $N = 3358$. Group 2 $N = 3363$. Group 3 $N = 3326$. Model 1 = 1-factor correlated traits model. Model 2 = 2-factor correlated traits model. Model 3 = 3-factor correlated traits model. BF = Bifactor Model. All values are significant at $p < 0.001$. [a] Because the robust χ^2 difference test resulted in a negative test statistic, a Wald Test was used to compare the 1- and 2-factor models for Kindergarten Group 1 and Grade 1 Group 2. The Wald tests for both Kindergarten in Group 1 (value = 83.15, $p < 0.001$) and Grade 1 in Group 2 (value = 59.83, $p < 0.001$) indicated that constraining the correlation between Inattention and Hyperactivity/impulsivity to 1.0 resulted in significantly poorer model fit

Overall Metric Invariance To examine whether the factors on the SWAN measured the same underlying behaviors across grades, metric invariance was evaluated by comparing models in which factor loadings were constrained to be equal across grades to models in which the factor loadings were freed for a specific grade level. In the model comparisons used to evaluate metric invariance, intercept values were estimated freely across grade level. Comparisons were examined using the chi-square difference test and change in CFI values. A summary of the results is shown in Table 3. Evidence of non-invariance was judged to be present if the lack of invariance was replicated in at least two of the three sub-samples for each comparison. As seen in Table 3, there was a consistent significant difference between the model with factor loadings constrained to equality and the

model with factor loadings freed for preschool across groups, although changes in the CFI value were minimal (i.e., Δ CFI = 0.001).

Construct-Level Metric Invariance For each comparison in which overall metric invariance was not supported (i.e., across the preschool groups), additional analyses were conducted at the factor (construct) level (e.g., inattention, H/I, ADHD) to determine the source of the lack of invariance. Results are shown in the lower panel of Table 3. Lack of metric invariance on the General ADHD Factor was found for preschool across groups. Both specific factors demonstrated metric invariance.

Item-Level Metric Invariance Wald Tests were conducted to examine the equality of factor loadings across items on the ADHD General Factors for preschool. Wald test values were

Table 3 Results of invariance testing presented by group

| | Overall Metric Invariance | | | | | | | | | | | | | | |
| | Group 1 | | | | | Group 2 | | | | | Group 3 | | | | |
Model	χ^2	scaling	$\Delta \chi^2$	CFI	Δ CFI	χ^2	scaling	$\Delta \chi^2$	CFI	Δ CFI	χ^2	scaling	$\Delta \chi^2$	CFI	Δ CFI
Overall metric invariance															
Fully Constrained	2851.33	1.58	–	0.956	–	2754.76	1.59	–	0.959	–	2521.37	1.63	–	0.963	–
PK released	2784.40	1.57	73.61[*]	0.957	0.001	2693.11	1.58	68.47[*]	0.960	0.001	2430.10	1.63	91.27[*]	0.964	0.001
K released	2811.67	1.57	50.02	.956	.000	2728.46	1.58	37.86	.959	.000	2499.27	1.63	22.10	.963	0.000
Grade 1 released	2783.42	1.58	67.91[*]	0.957	0.001	2712.65	1.59	42.11	.959	0.000	2491.72	1.63	29.65	.963	0.000
Grade 2 released	2819.47	1.57	43.27	.956	0.000	2702.76	1.59	52.00	0.960	0.001	2491.68	1.63	29.69	.963	0.000
Grade 3 released	2795.08	1.58	56.25	.957	.001	2725.83	1.60	13.83	.959	0.000	2493.95	1.63	27.42	.963	0.000
Grade 4 released	2804.81	1.58	46.52	.957	.001	2717.30	1.58	47.53	0.959	0.000	2446.06	1.63	75.31[*]	0.964	0.001
Overall scalar invariance															
Metric	2871.50	1.54	–	.956		2844.17	1.55		0.958		2482.80	1.60		0.963	
PK released	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
K released	2788.69	1.55	93.12[*]	0.957	0.001	2808.14	1.55	36.03[*]	0.959	.001	2429.59	1.61	53.84[*]	0.964	0.001
Grade 1 released	–	–	–	–	–	2810.69	1.55	33.48	.959	.001	2465.23	1.61	3.06	0.963	0.000
Grade 2 released	2841.84	1.55	16.13	.956	0.000	2814.21	1.55	29.96	.958	.000	2432.47	1.61	49.74[*]	0.964	0.001
Grade 3 released	2826.63	1.55	38.16[*]	0.956	0.000	2806.32	1.55	37.85[*]	0.959	0.001	2445.87	1.61	30.64	0.964	0.001
Grade 4 released	2830.10	1.55	33.13	.956	.000	2795.12	1.55	49.05[*]	0.959	0.001	–	–	–	–	–
Construct level metric invariance															
Fully Constrained	2851.33	1.58	–	0.956	–	2754.76	1.59	–	0.959	–	2521.37	1.63	–	0.963	–
PK															
General released	2799.43	1.58	51.90[*]	0.957	0.001	2697.63	1.60	57.53[*]	0.960	0.001	2460.44	1.64	64.97[*]	0.964	0.001
IA released	2846.06	1.57	23.22	.956	0.000	2735.54	1.59	19.22	.960	.001	2496.41	1.63	24.96	.963	0.000
HI released	2841.32	1.57	27.91	.956	0.000	2769.76	1.58	2.41	.960	.001	2514.16	1.63	7.21	.963	0.000

Group 1 N = (PK = 503; Grade 1 = 612). Group 2 N = (PK = 528; Grade 1 = 647; Grade 2 = 520). Group 3 N = (PK = 501; Grade 1 = 657; Grade 2 = 553; Grade 4 = 420). Bolded values are significant at *p* < 0.003. Factor loadings were constrained to equality across grades in the base model for metric invariance testing. DF = 882 in the base model for metric invariance testing. DF = 846 for models in which factor loadings are released for a single grade. DF = 864, 918, and 864 in the base model for scalar invariance testing in groups 1, 2, and 3, respectively. DF = 846, 900, and 846 for models in which item intercepts are released for a single grade for group 1, 2, and 3, respectively. DF = 864, 873, 873 for models in which the factor loadings were released for general factor, the IA factor, and the HI factor, respectively. * Sig at < 0.008. Scalar invariance was only tested for groups which were found to have metric invariance

considered significant at $p \leq 0.003$. Overall, there very few patterns across groups. One inattention item (i.e., "Sustain attention on tasks or play activities") was found to be non-invariant in all groups (Group 1 Wald test value = 11.67; Group 2 Wald test value = 19.67; Group 3 Wald test value = 10.94). Four inattention items (i.e., "Organizes tasks or activities," "Engages in tasks that require sustained mental effort," "Keeps track of things necessary for activities," and "Ignores extraneous stimuli") were found to be non-invariant across two groups (Wald test values between 8.52 and 22.23). Three hyperactivity items (i.e., "Play quietly [keep noise level reasonable]," "Settles down and rests [controls constant activity]," "Modulate verbal activity [control excess talking]") and one impulsivity item (i.e., "Reflects on questions [controls blurting out answers]") were found to be non-invariant across two groups (Wald test values between 8.70 and 16.74).

The unstandardized factor loadings for each item, when the loadings for the elementary grades are held to be equal, listed separately for preschool and the elementary grade levels, are presented in Table 4. The standardized factor loadings derived from the models are available in section C of the supplemental materials. Across groups and items, factor loadings on the general ADHD factor were, with few exceptions, higher for the invariant grades than they were for preschool.

Scalar Invariance To examine whether the means of specific items were similar across grades at the mean levels of the latent constructs, scalar measurement invariance was evaluated by comparing models in which item intercepts were constrained to be equal across grades to models in which the item intercepts were freed for a specific grade level. Across groups, scalar invariance was only evaluated for grades in which metric invariance was established. Comparisons were examined using the chi-square difference test and change in CFI values. A summary of the results is shown in Table 3. Overall, only Grade 1 and Grade 2 demonstrated scalar invariance in at least two of the groups. Modification indices were examined to explore whether specific items were driving the lack of invariance across grades and groups. A modification index value of 3.84 (i.e., $p < 0.001$) was used to examine items that, if freed from constraint across grades, would result in significant improvement to model fit. In Group 1, modification indices were significant for items 1, 3, 8, 12, 13, and 14 in Kindergarten and for items 1 and 4 in Grade 3. In Group 2, modification indices were significant for item 12 in Kindergarten, items 1 and 3 in Grade 1, for items 1 and 3 in Grade 3, and item 9 in Grade 4. In Group 3, modification indices were significant for items 4 and 13 in Kindergarten, for items 10, 12, and 13 in Grade 2 and for items 3 and 9 in Grade 3. In sum, no consistent patterns emerged across groups or grades.

Table 4 Factor Loadings in the Bifactor Model (with Elementary Grades Combined)

| Item | Group 1 | | | | Group 2 | | | | Group 3 | | | |
| | Preschool | | Elementary | | Preschool | | Elementary | | Preschool | | Elementary | |
	S	G	S	G	S	G	S	G	S	G	S	G
1	**0.86**	**0.88**	**0.89**	**1.03**	**0.88**	**0.84**	**0.96**	**1.10**	**0.83**	**0.97**	**0.94**	**1.08**
2	**0.82**	**0.87**	**0.81**	**1.10**	**0.83**	**0.81**	**0.87**	**1.15**	**0.72**	**0.98**	**0.82**	**1.16**
3	**0.59**	**1.00**	**0.59**	**1.02**	**0.47**	**0.94**	**0.63**	**1.11**	**0.52**	**0.99**	**0.62**	**1.07**
4	**0.79**	**1.04**	**0.92**	**1.11**	**0.79**	**0.94**	**0.98**	**1.18**	**0.71**	**1.08**	**0.95**	**1.16**
5	**0.83**	**0.93**	**0.91**	**1.05**	**0.81**	**0.84**	**0.98**	**1.11**	**0.76**	**0.94**	**0.95**	**1.11**
6	**0.86**	**0.96**	**0.98**	**1.03**	**0.96**	**0.79**	**1.02**	**1.11**	**0.88**	**0.87**	**0.98**	**1.13**
7	**0.81**	**0.91**	**0.84**	**1.05**	**0.80**	**0.83**	**0.90**	**1.13**	**0.76**	**0.90**	**0.90**	**1.11**
8	**0.50**	**0.94**	**0.54**	**1.10**	**0.43**	**0.85**	**0.57**	**1.21**	**0.37**	**0.91**	**0.55**	**1.18**
9	**0.64**	**0.85**	**0.70**	**0.90**	**0.65**	**0.79**	**0.74**	**0.99**	**0.60**	**0.87**	**0.73**	**0.97**
10	**0.45**	**1.29**	**−0.21**	**1.39**	**−0.32**	**1.16**	**−0.33**	**1.46**	**−0.20**	**1.28**	**−0.25**	**1.42**
11	**0.34**	**1.29**	**−0.11**	**1.37**	**−0.38**	**1.21**	**−0.26**	**1.44**	**−0.23**	**1.24**	**−0.17**	**1.4**
12	0.07	**1.02**	0.07	**1.02**	−0.06	**0.83**	−0.03	**1.13**	0.05	**0.98**	0.03	**1.07**
13	0.07	**1.20**	**0.32**	**1.18**	0.13	**1.04**	**0.18**	**1.28**	**0.24**	**1.04**	**0.26**	**1.21**
14	−0.06	**1.11**	**0.23**	**1.23**	0.04	**1.04**	0.07	**1.32**	**0.15**	**1.04**	**0.14**	**1.27**
15	−0.05	**1.15**	**0.43**	**1.25**	**0.33**	**1.04**	**0.30**	**1.35**	**0.30**	**1.06**	**0.37**	**1.28**
16	**−0.23**	**1.10**	**0.49**	**1.19**	**0.43**	**1.07**	**0.40**	**1.32**	**0.41**	**1.08**	**0.43**	**1.30**
17	**−0.22**	**1.11**	**0.38**	**1.13**	**0.29**	**1.09**	**0.35**	**1.26**	**0.28**	**1.10**	**0.36**	**1.21**
18	**−0.26**	**1.09**	**0.39**	**1.06**	**0.35**	**0.97**	**0.34**	**1.20**	**0.44**	**1.02**	**0.35**	**1.14**

S = Factor loading on the respective specific factor; G = Factor Loading on the general factor. Bolded values are significant at $p < 0.003$

Discussion

The two goals of this study were to examine the structure of inattentive, hyperactive, and impulsive behaviors in a large sample of children ranging from preschool to grade 4 and to examine whether the measurements of these behavioral constructs were consistent across grade levels. A bifactor model with a general ADHD factor and two specific Inattention and H/I factors provided the best fit to the data. Results supporting the bifactor model were consistent across all grades and for each of three subgroups of children within each grade. Overall, item loadings were moderate to high for the general ADHD factor and for the specific Inattention factor. Item loadings, particularly for the hyperactivity items, were low and sometimes inconsistent across subsamples for the specific H/I factor. Together, the mixed findings regarding the distinction of hyperactivity and impulsivity and the inconsistent item-loadings on the specific H/I factor challenge the conceptualization of hyperactivity and impulsivity as a unitary presentation of ADHD-related behaviors. Further, the weak loadings on the specific factor, despite strong loadings on the general factors suggest that hyperactive/impulsive behaviors may be implicated in halo effects that contribute to the high interrelatedness among teachers' ratings of the different behavioral dimensions associated with ADHD (Hartung et al. 2010). Still, the consistent finding that hyperactivity and impulsivity fit the data well when modeled as a single factor, suggests that these constructs are best conceptualized as unidimensional. Thus, findings regarding the dimensionality of hyperactivity and impulsivity were mixed.

In general, the bifactor model demonstrated metric invariance from kindergarten through grade 4 but not for preschool. The results of this study regarding the structure of the SWAN support the conceptualization of ADHD as a single disorder with different sub-presentations. The finding of metric invariance from kindergarten to grade 4 indicate that inattention, H/I, and the variance common among these behaviors reflect the same underlying constructs across this age range. The lack of metric invariance between preschool and the other grade levels indicates that the problem behaviors associated with general ADHD differ for preschool children.

Structure of Inattentive, Hyperactive, and Impulsive Behaviors

Tests examining the factor structure of inattention and H/I indicated that inattentive, impulsive, and hyperactive behaviors were best conceptualized as behavior constructs with both unitary and distinct features. The finding that a bifactor model comprising a General ADHD Factor and two specific factors provided the best fit to the model across all grades and all group suggests that this is a robust conceptualization of inattentive, hyperactive, and impulsive behaviors. The emergence of a general underlying ADHD factor is in line with other recent analyses of inattentive and hyperactive/impulsive behaviors using bifactor and hierarchical modeling techniques (e.g., Toplak et al. 2012; Ullebø et al. 2012; Willoughby et al. 2015). This common finding across studies offers further support for grouping these heterogeneous behavioral constructs under the classification of a single disorder.

The finding that the factor loadings for the items on the specific H/I factor were often weak and sometimes negative suggests that although there may be unique aspects of inattention that are distinct from more general problem behaviors associated with ADHD, behaviors characterizing H/I may be better understood as reflective of the more general problem behaviors associated with ADHD. The pattern of frequent negative loadings on the specific H/I factor likely represents a suppression effect in which the general ADHD factor accounts for such a large amount of variance in items that the relations between the items and the remaining variance left over to be accounted for by the specific factor were inflated and related in the opposite direction than would be expected.

The unexpected results regarding the specific H/I Factor mirror unexpected results reported by other researchers who have modeled specific hyperactivity or specific combined H/I factors. For example, using bifactor modeling and symptom classifications based on the ICD-10, Ullebø et al. (2012) reported support for a general ADHD factor and two specific Inattention and Impulsivity factors, but their results did not support a specific Hyperactivity factor. Willoughby et al. (2015) reported results supporting a bifactor model with a general ADHD factor and two specific Inattention and combined H/I factors. In their study, Willoughby et al. reported that the specific H/I factor had few significant associations with other outcomes, such as academic skills and treatment utilization. Although the specific H/I factor was associated with aspects of social development, it was related opposite to the expected direction (e.g., greater hyperactivity was associated with more closeness to teachers). Thus, the empirical analyses of the variance in ADHD-related behaviors conducted in the present study and previous research suggests that ADHD rating scales based on the 18 symptoms of ADHD may not adequately capture unique and meaningful specific aspects of hyperactivity and impulsivity.

The presence of a well-defined General ADHD factor supports the conceptualization of ADHD, despite its sub-presentations, as a unified disorder. However, the degree of shared variance among ratings of individual items may also reflect potential halo effects. Halo effects refer to potential biases in informant ratings that occur when an individual's general impressions of a child unduly influence her or his evaluations of the child. In the context of behavior-ratings, halo effects may lead informants who observe one type of behavior to report the presence of another, related behavior that was actually not observed (e.g., Hartung et al. 2010). The finding that the hyperactivity and impulsivity items on the SWAN demonstrated

strong loadings on the general ADHD factor but failed to form a strong and consistent specific factor suggests that hyperactive/impulsive behaviors may most heavily influence possible halo effects. It may be that these behaviors are the most readily observable, leading an informant who observes these hyperactive and impulsive behaviors to draw inferences regarding other less noticeable behaviors such as those associated with inattention.

Similar to the extant research on the distinctness of hyperactivity and impulsivity (see Willcutt et al. 2012 for review), the results of this study raise several mixed points regarding the separation of these two closely related constructs. Taken together, the strong correlation between hyperactivity and impulsivity and the poor reliability of the impulsivity factor in the 3-factor correlated traits model suggested limited utility in separating these closely related behaviors. As such, the bifactor model tested in the present study was based on the 2-factor correlated traits model. However, it is notable that changes in both the chi-square and approximate fit indices pointed to the 3-factor correlated traits as the best fit to the data. This indicates that despite the strong interrelatedness of these constructs, modeling hyperactivity and impulsivity as distinct constructs results in empirical improvements to the modeling of variance in ADHD-related behaviors. This finding echoes concerns discussed by others (e.g., Coghill and Seth 2011; Frick and Nigg 2012) that impulsivity may deserve further consideration as a distinct aspect of ADHD. Given the influence of scale size on reliability estimates (Trizano-Hermosilla and Alvarado 2016), it may be that more items are needed for impulsivity scales to reflect a reliable unitary construct with variance that allows it to be distinguished empirically from hyperactivity. Although many discussions concerning the addition of impulsivity items have focused on adults (e.g., Coghill and Seth 2011; Frick and Nigg 2012), our findings suggest that this issue also is relevant in early and middle childhood populations.

It may be that hyperactivity and impulsivity are indeed distinct constructs; however, the overlap in the functional impairment caused by these difficulties and possible shared cognitive mechanisms that lead to the expression of both hyperactivity and impulsivity make it difficult to separate these dimensions (Raiker et al. 2012). Research is needed to determine whether there are meaningful differential relations between these two types of problem behaviors, in the context of a bifactor model, and other important outcomes such as social skills, academic skills, and other developmental disorders. A more nuanced understanding of the distinctness of these constructs may guide the field's understanding of the etiology of different manifestations of ADHD and may improve how the disorder is conceptualized, diagnosed, and treated.

Measurement Invariance

Results supported metric invariance for inattentive and hyperactive/impulsive behaviors from kindergarten through fourth-grade. Although children undergo substantial cognitive, emotional, and behavioral changes from kindergarten to fourth-grade, this finding indicates that similar behaviors reflect underlying inattention, H/I, and general ADHD behaviors across this developmental span. In contrast to results across the elementary school grades, measurement of the behaviors associated with ADHD differed between preschool and the elementary school grades. In general, factor loadings were smaller in preschool than they were for the older grades. More detailed analysis at the construct level revealed that the lack of invariance was attributable to differences between children in preschool and children in other grades in how items on the SWAN related to the General ADHD factor and not how the items on the SWAN related to the specific Inattention and H/I factors. These results indicate that behaviors associated with the unique aspects of inattention and H/I are generally similar across development from preschool to fourth grade; however, how these behaviors reflect the shared variance of inattention, hyperactivity, and impulsivity differs between the preschool and elementary school years.

There are several explanations for the differences in the measurement of general ADHD behaviors in preschool compared to other grade levels. It could be that there are genuine developmental differences between preschool children and children in the primary grades. The executive processes used to control inattention, hyperactivity, and impulsivity are rapidly developing during the preschool years (Espy 2004). Inattentive, hyperactive, and impulsive behaviors are relatively common in preschool-age children, and, thus, the presence of these behaviors may not indicate overall behavior problems in preschool children to the same extent they do in older children. The general variability in preschooler's behavior may attenuate the precision of the measurement of problem behaviors, which might explain the generally lower factor loadings for this age-group.

Differences in how behaviors characterized by inattention, hyperactivity, and impulsivity relate to a common ADHD construct also may be the result of differences in teacher characteristics, teacher expectations, and setting demands in preschool compared to older grades. Whereas teaching positions in elementary school generally require a post-secondary degree, degree requirements for teaching positions in preschool are more varied. Preschools teachers' educational backgrounds may range from masters level degree to less than a high school degree, depending on the organization of which the preschool is a part. As such, preschool teachers vary in terms of their training, their views on classroom structure, and their expectations for children's roles and behaviors in the classroom.

Many of the behaviors used to operationalize the presence of inattention and H/I require teachers to observe the ability, or lack thereof, to perform tasks that may or may not be expected of children in a preschool classroom. The item that was non-invariant between preschool and older grades across all three groups refers to the ability to "sustain" attention. Given that activities in preschool classrooms are typically more play-based, fast moving, and loosely structured than are activities in the primary grades, preschool teachers may have fewer opportunities to observe children in "sustained" activities. Other inattention items that were non-invariant across two groups (e.g., "Organizes tasks and activities," "Engage in tasks that require sustained mental effort," "keep track of things necessary for activities") involved activities that may be less relevant in preschool classrooms. Consequently, the same behaviors in preschool and elementary grades may reflect different degrees of overall difficulties. Similarly, the hyperactivity items that were not invariant in at least two groups (i.e., "play quietly," "settle down," and "modulate verbal activity.") may reflect behaviors that are less important or salient in less-structured and play-based preschool classrooms than they are in elementary-level grades. Similarly, the impulsivity item that was non-invariant between preschool and older grades in two groups was "reflects on questions (controls blurting out answer)." Compared to Elementary school, whole-group activities in preschool tend to be less formal and less dependent on "correct or incorrect" answers. Therefore, judging a child's ability to "reflect" in preschool may be more difficult than judging other impulsive behaviors such as awaiting turns and entering games without intruding.

In summary, the underlying mechanism driving the differences in the measurement of inattentive, hyperactive, and impulsive behaviors between preschool children and older children are not completely clear. Regardless, the results of this study suggest that caution should be taken when interpreting preschoolers' scores on measures of inattention and H/I that are typically used with populations of older children. Further, the same behaviors observed in preschool and older grades may have differential impacts on teachers' conceptualizations of children's overall levels of problem behaviors.

Limitations and Future Directions

There are several strengths of this study including the use of a large sample of both preschool and school-age children and the use of a relatively novel statistical technique that may be particularly useful for understanding the complex structure of the behaviors associated with ADHD; however, it is not without limitations. For one, this study was conducted using a community-based sample rather than a sample of children who have been identified as exhibiting problem behaviors. Although a range of scores on the SWAN, including scores in the clinical range, were present in this sample, it is possible

that the structure and measurement of inattention, hyperactivity, and impulsivity would differ in a sample of mostly children with clinical levels of these behaviors. As such, the findings of the present study only generalize to the structure and measurement of inattentive and hyperactive/impulsive behaviors as they occur in the general population. Research on a clinical sample would be needed to establish whether these results generalize to children with diagnosed ADHD. Further, although the use of a bifactor model represents a significant strength of this study, no external variables were included to examine whether the general and specific factors had differential and meaningful relations to other childhood outcomes (e.g., academic skills, social skills).

Another limitation was the use of a single method (i.e., teacher-ratings) of measuring inattention, hyperactivity, and impulsivity. Further these ratings were made after the teachers had interacted with the child for no more than two or three months. Although teachers are considered ideal informants of children's behavior (e.g., Evans et al. 2005), parents are another important informant of their child's behaviors. However, parent-ratings were not available for the majority of children in the present study. It is possible that the results would have differed for ratings provided by other informants (e.g., parents, self). Thus, a comprehensive understanding of the structure and measurement of inattention and H/I will require that the factor structure and measurement invariance of inattention and H/I be examined using ratings provided by other informants.

Given the aforementioned differences between preschool and elementary-school level classrooms and teachers, it is possible that school-related factors contributed to the observed lack of invariance across preschool and school-age children. To address the possibility that classroom expectations drive differences in measurement, future research should examine whether the measurement of inattentive and hyperactive/impulsive behaviors differs across preschools with varying structures (e.g., those that primarily focus on free-play compared to those with a direct-instruction component).

Conclusions

The results of this study indicate that inattention, hyperactivity, and impulsivity were best characterized as behavioral constructs with both shared and unique components. The lack of strong specific H/I factor challenges the utility of conceptualizing H/I, as rated by teachers, as a behavioral construct that is distinct from the general problem behaviors that are accounted for by the shared variance among all three ADHD-related behaviors. Although the practice of merging hyperactive and impulsive behaviors for diagnostic purposes persists, the findings of the current study suggest that further examination of the distinct aspects of these behavioral constructs is warranted.

The findings of this study related to measurement invariance testing point to the need for caution when interpreting differences in inattention and H/I between preschoolers and older children. Informants' ratings of inattention and H/I may not necessarily provide an assessment of the same underlying behavioral constructs across early and middle childhood. Further inquiry into the underlying reasons for the observed differences between preschooler and school-age children (e.g., developmental, measurement-related) may help guide and improve the early identification of children who are at-risk for developing long-term behavioral difficulties.

Compliance with Ethical Standards

Conflict of Interest   The authors have no potential conflict of interest to report.

Ethical Approval   Procedures for all projects were approved by Florida State University's Institutional Review Board.

Informed Consent   Informed consent was obtained from children's parents or legal guardians before data collection began.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM- IV* (4th 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.

Arias, V. B., Ponce, F. P., Martínez-Molina, A., Arias, B., & Núñez, D. (2016). General and specific attention-deficit/hyperactivity disorder factors of children 4 to 6 years of age: An exploratory structural equation modeling approach to assessing symptom multidimensionality. *Journal of Abnormal Psychology, 123*, 125–137. https://doi.org/10.1037/abn0000115.

Arnett, A. B., Pennington, B. F., Friend, A., Willcutt, E. G., Byrne, B., Samuelsson, S., & Olson, R. K. (2013). The SWAN captures variance at the negative and positive ends of the ADHD symptom dimension. *Journal of Attention Disorders, 17*, 152–162.

Barbaresi, W. J., Colligan, R. C., Weaver, A. L., Voigt, R. G., Killian, J. M., & Katusic, S. K. (2013). Mortality, ADHD, and psychosocial adversity in adults with childhood ADHD: A prospective study. *Pediatrics, 131*, 637–644. https://doi.org/10.1542/peds.2012-2354.

Biederman, J., Mick, E., & Faraone, S. V. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: Impact of remission definition and symptom type. *American Journal of Psychiatry, 5*, 816–818.

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. https://doi.org/10.1023/A:1010314030025.

Caci, H. M., Morin, A. J., & Tran, A. (2016). Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders, 20*, 434–444. https://doi.org/10.1177/1087054712473834.

Coghill, D., & Seth, S. (2011). Do the diagnostic criteria for ADHD need to change? Comments on the preliminary proposals of the DSM-5 ADHD and disruptive behavior disorders committee. *European Child & Adolescent Psychiatry, 20*, 75–81. https://doi.org/10.1007/s00787-010-0142-4.

Connor, C. M., Phillips, B. M., Kim, Y.-S. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & Al Otaiba, S. (2018). Examining the efficacy of targeted component interventions on language and literacy for third and fourth graders who are at risk of comprehension difficulties. *Scientific Studies of Reading, 22*, 462–484. https://doi.org/10.1080/10888438.2018.1481409.

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. https://doi.org/10.1037/1040-3590.9.4.436.

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*, 214–225. https://doi.org/10.1037/pas0000166.

Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. *Developmental Neuropsychology, 26*, 379–384. https://doi.org/10.1207/s15326942dn2601_1.

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

Frick, P. J., & Nigg, J. T. (2012). Current issues in the diagnosis of attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. *Annual Review of Clinical Psychology, 8*, 77–107. https://doi.org/10.1146/annurev-clinpsy-032511-143150.

Gómez, R., Vance, A., & Gómez, R. M. (2017). Validity of the ADHD bifactor model in general community samples of adolescents and adults, and a clinic-referred sample of children and adolescents. *Journal of Attention Disorders.* https://doi.org/10.1177/1087054713480034.

Hartung, C. M., Lefler, E. K., Tempel, A. B., Armendariz, M. L., Sigel, B. A., & Little, C. S. (2010). Halo effects in ratings of ADHD and ODD: Identification of susceptible symptoms. *Journal of Psychopathology and Behavioral Assessment, 32*, 128–137. https://doi.org/10.1007/s10862-009-9135-3.

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

Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry, 62*, 896–902. https://doi.org/10.1001/archpsyc.62.8.896.

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The reliability and validity of the English and Spanish strengths and weaknesses of ADHD and Normal behavior rating scales in a preschool sample continuum measures of hyperactivity and inattention. *Journal of Attention Disorders, 16*, 510–516. https://doi.org/10.1177/1087054711413550.

Larsson, H., Dilshad, R., Lichtenstein, P., & Barker, E. D. (2011). Developmental trajectories of DSM-IV symptoms of attention-deficit/hyperactivity disorder: Genetic effects, family risk and associated psychopathology. *Journal of Child Psychology and Psychiatry, 52*, 954–963. https://doi.org/10.1111/j.1469-7610.2011.02379.x.

Leopold, D. R., Christopher, M. E., Burns, G. L., Becker, S. P., Olson, R. K., & Willcutt, E. G. (2016). Attention-deficit/hyperactivity disorder and sluggish cognitive tempo throughout childhood: Temporal invariance and stability from preschool through ninth grade. *Journal of Child Psychology and Psychiatry, 57*, 1066–1074. https://doi.org/10.1111/jcpp.12505.

Lonigan, C. J., & Burgess, S. R. (2017). Dimensionality of reading skills with elementary-school-age children. *Scientific Studies of Reading, 21*, 239–253. https://doi.org/10.1080/10888438.2017.1285918.

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of typically developing preschool through fifth-grade children. *Journal of Speech, Language, and Hearing Research, 60*, 2185–2198. https://doi.org/10.1044/2017_JSLHR-L-15-0402.

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the simple view of reading with elementary school children: Still simple after all these years. *Remedial and Special Education, 39*, 260–273. https://doi.org/10.1177/0741932518764833.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130.

Martel, M. M., Roberts, B., Gremillion, M., von Eye, A., & Nigg, J. T. (2011). External validation of bifactor model of ADHD: Explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *Journal of Abnormal Child Psychology, 39*, 1111–1123. https://doi.org/10.1007/s10802-011-9538-y.

McGoey, K. E., Schreiber, J., Venesky, L., Westwood, W., McGuirk, L., & Schaffner, K. (2015). Factor structure of attention deficit hyperactivity disorder symptoms for children age 3 to 5 years. *Journal of Psychoeducational Assessment, 33*, 430–438. https://doi.org/10.1177/0734282914554255.

Molina, B. S., 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. https://doi.org/10.1023/A:1005203629968.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles: Muthén & Muthén.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 216–316 Retrieved from http://www.jstor.org/stable/271070.

Raiker, J. S., Rapport, M. D., Kofler, M. J., & Sarver, D. E. (2012). Objectively-measured impulsivity and attention-deficit/hyperactivity disorder (ADHD): Testing competing predictions from the working memory and behavioral inhibition models of ADHD. *Journal of Abnormal Child Psychology, 40*, 699–713.

Swanson, J. M., Schuck, S., Mann, M, Carlson, C., Hartman, K., Sergeant, J., et al. (2001). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN rating scales. Available at: http://adhd.net. Accessed 15 Nov 2015.

Toplak, M. E., Pitch, A., Flora, D. B., Iwenofu, L., Ghelani, K., Jain, U., & Tannock, R. (2009). The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: Evidence for a general factor with separable dimensions. *Journal of Abnormal Child Psychology, 37*, 1137–1150. https://doi.org/10.1007/s10802-009-9336-y.

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

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*(769). https://doi.org/10.3389/fpsyg.2016.00769.

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. https://doi.org/10.1111/j.1469-7610.2012.02549.x.

Wagner, F., Martel, M. M., Cogo-Moreira, H., Maia, C. R. M., Pan, P. M., Rohde, L. A., & Salum, G. A. (2015). Attention-deficit/hyperactivity disorder dimensionality: The reliable 'g' and the elusive 's' dimensions. *European Child & Adolescent Psychiatry, 25*, 1–8. https://doi.org/10.1007/s00787-015-0709-1.

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

Willoughby, M. T., Blanton, Z. E., & Investigators, F. L. P. (2015). Replication and external validation of a bi-factor parameterization of attention deficit/hyperactivity symptomatology. *Journal of Clinical Child & Adolescent Psychology, 44*(1), 68–79.

Wolraich, M. L., Lambert, E. W., Baumgaertel, A., Garcia-Tornel, S., Feurer, I. D., Bickman, L., & Doffing, M. A. (2003). Teachers' screening for attention deficit/hyperactivity disorder: Comparing multinational samples on teacher ratings of ADHD. *Journal of Abnormal Child Psychology, 31*, 445–455. https://doi.org/10.1023/A:1023847719796.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*, 165–200. https://doi.org/10.1111/0081-1780.00078.