Rethinking hyperactivity in pediatric ADHD: Preliminary evidence for a re-conceptualization of hyperactivity/impulsivity from the perspective of informant perceptual processes

Michael J. Kofler, Ph.D.[1], **Nicole B. Groves, M.S.**[1], **Leah J. Singh, Ph.D.**[1], **Elia F. Soto, M.S.**[1], **Elizabeth S.M. Chan, M.A.**[1], **Lauren N. Irwin, M.S.**[1], **Caroline E. Miller, B.A.**[1]

[1]Florida State University, Department of Psychology

Abstract

Hyperactivity is a core ADHD symptom that has been both positively and negatively associated with cognition and functional outcomes. The reason for these conflicting findings is unclear but may relate to subjective assessments that conflate excess physical movement (hyperactivity) with verbally intrusive/impulsive behaviors. The current study adopted a model-driven, rational-empirical approach to distinguish excess physical movement symptoms from other, auditorily-perceived behaviors assessed under the ‘hyperactivity/impulsivity’ umbrella. We then tested this alternative conceptualization’s fit, reliability, replicability, convergent/divergent validity via actigraphy, and generalizability across informants (parents, teachers) in a well-characterized, clinically-evaluated sample of 132 children ages 8–13 years (*M*=10.34, *SD*=1.51; 47 girls; 67% White/non-Hispanic). The current DSM hyperactivity/impulsivity item pool can be reliably reclassified by knowledgeable judges into items reflecting excess physical movement (visual hyperactivity) and auditory interruptions (verbal intrusion). This bifactor structure showed evidence for multidimensionality and superior model fit relative to traditional hyperactivity/impulsivity models. The resultant visual hyperactivity factor was reliable, replicable, and showed strong convergent validity evidence via associations with objectively-assessed hyperactivity. The verbal intrusion factor also showed evidence for reliability and explained a substantive portion of reliable variance, but demonstrated lower estimated replicability. These findings provide preliminary support for conceptualizing ADHD symptoms from the perspective of their cognitive-perceptual impact on others, as well as differentiating excess physical movement (hyperactivity) from other behaviors assessed under the ‘hyperactivity/impulsivity’ umbrella. ‘Verbal intrusion’ appears to provide a better explanation than ‘impulsivity’ for the reliable, non-hyperactivity variance assessed by these items, but the current item set appears insufficient for replicable measurement of this construct.

Corresponding Author: Michael J. Kofler, Ph.D., Florida State University | Department of Psychology, 1107 W. Call Street | Tallahassee, FL 32306-4301, Phone: (850) 645-0656, Fax: (850) 644-7739, kofler@psy.fsu.edu.

Conflict of Interest: The authors have no conflicts of interest to report.

Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Keywords

ADHD; hyperactivity; verbal intrusion; impulsivity; actigraph; working memory; bifactor

Attention-deficit/hyperactivity disorder (ADHD) is a chronic and impairing neurodevelopmental disorder that affects 5% of school-aged children (Polanczyk et al., 2007, 2014) at an annual U.S. cost of illness of over $100 billion (Zhao et al., 2019). Hyperactivity, or excess physical movement, was considered the disorder's dominant clinical feature throughout most of the 20th century before being relegated to a secondary role and lumped with impulsivity symptoms when the DSM-III was published in 1980 (for reviews, see Martinez-Badia & Martinez-Raga, 2015; Rapport et al., 2009). Although the behavioral indicators have changed somewhat over time, excess physical movement (hyperactivity) and behaviors interpreted by others as 'acting without thinking' (impulsivity) have remain grouped together under the hyperactivity/impulsivity label across DSM-IV and DSM-5 revisions (APA, 2013). Interestingly, hyperactivity is considered a core and impairing ADHD symptom domain despite a surprising number of studies reporting *positive* associations with cognition and important functional outcomes (e.g., Kofler et al., 2018; Sarver et al., 2015). As a first step toward synthesizing these discrepant findings, the current study applies cognitive models of perception and memory (Bettencourt & Xu, 2016) to the study of overt behavior and examines the feasibility and utility of maximally differentiating excess physical movement from other behaviors assessed under the 'hyperactivity/impulsivity' umbrella.

Hyperactivity in ADHD: Compensatory behavior or impairing deficit?

Replicated evidence indicates that children with ADHD perform *better* on challenging neurocognitive tasks when they are more physically active (Hartanto et al., 2016; Sarver et al., 2015). Further, experimental and meta-analytic evidence indicates a strong positive association between excess physical movement and cognition, such that children with and without ADHD show large increases in physical activity during cognitively challenging activities (Hudec et al., 2015; Kofler et al., 2016, 2018; Patros et al., 2017; Rapport et al., 2009). In addition, increased hyperactivity symptoms have been linked with *better* teacher-reported organizational skills associated with task planning in children with ADHD (Kofler et al., 2018). Evidence suggestive of a causal role of hyperactivity for facilitating cognitive and functional outcomes in children with ADHD comes from studies that have experimentally induced higher levels of physical activity and demonstrated clinically significant improvements in cognitive (Chang et al., 2012; Gapin et al., 2011; Pontifex et al., 2015; Smith et al., 2013; Verret et al., 2012; Medina et al., 2010) and academic test performance (Pontifex et al., 2013) as well as informant perceptions of classroom deportment and peer interactions (Ahmed & Mohamed, 2011; Smith et al., 2013; Verret et al., 2012).

In contrast, a large body of evidence suggests negative associations between hyperactivity and a broad range of ADHD-related symptoms and functional impairments ranging from social problems (e.g., Bunford et al., 2014; Kofler, Harmon et al., 2018) to neurocognitive test performance (e.g., Brocki et al., 2010; Nigg et al., 2002). In addition, the prevailing DSM clinical model conceptualizes hyperactivity as a core, impairing deficit (APA, 2013), and empirical evidence indicates that parents and teachers explicitly link children's functional impairments with their hyperactivity/impulsivity symptoms to the same degree seen for their inattention symptoms (DuPaul et al., 2016).

How, then, can we resolve this apparent discrepancy in which hyperactivity is classified as both a useful compensatory behavior and an impairing symptom across different studies? Two possibilities merit scrutiny. First, the discrepant findings may reflect differences in measurement and methodology, such that most studies finding positive hyperactivity/outcome associations used mechanical measures of excess physical movement (e.g., actigraphs) and within-subject designs that evaluated children's performance relative to themselves (e.g., Sarver et al. 2015). Thus, these studies were able to objectively and specifically assess hyperactivity without the potential confounds associated with informant perception and assumptions regarding the movement's intent.

In contrast, most studies finding negative associations used between-subject designs and informant ratings that may conflate physical movement with other symptoms (e.g., impulsivity). To that end, the ability to reliably and specifically assess physical movement is critical for intervention planning. For example, to the extent that physical movement augments physiological arousal and improves cognition in children with ADHD, behavioral approaches that overly restrict physical movement may have unintended, adverse effects on cognitive and academic functioning as argued previously (Rapport et al., 2009). This line of reasoning has led to the growing popularity of 'squirm to learn' or 'freedom to move' approaches to ADHD in the classroom, which allow higher levels of physical movement while still targeting verbal intrusions and other behaviors that are disruptive to peers (Kofler et al., 2016). However, the efficacy of these approaches remains unknown, in part due to measurement challenges associated with differentiating excess motor movement from other behaviors assessed under the 'hyperactivity/impulsivity' umbrella as described below.

Sensory perception and processing: Modality-specific adverse effects on others

A related explanation that warrants scrutiny is the lumping of items intended to assess excess physical movement (e.g., fidgeting, getting out of seat) with items that may assess conceptually and functionally distinct behaviors (e.g., excessive talking, blurting out). This possibility relates to the use of subjective, informant-based ratings that may obscure hyperactivity's proposed compensatory effects (Rapport et al., 2009) secondary to the reliance on behaviorally-anchored items that may load together as an artifact of insufficiently capturing the intent (or lack thereof) of a behavior (De Los Reyes & Kazdin, 2005). For example, contrary to DSM-based conceptualizations of hyperactivity and impulsivity as reflective indicators of a single symptom domain (APA, 2013), trait impulsivity is often parsed into distinct constructs that have been differentially linked with a wide range of personality, behavioral, and neurocognitive correlates (Sharma et al., 2014). Further, most measures of trait impulsivity exclude or only minimally assess behaviors associated with physical/motor movement (Whiteside & Lynam, 2001), with impulsivity/activity level associations interpreted as evidence for 'functional' forms of impulsivity (Dickman, 1990) – a distinction conceptually similar to recent distinctions between compensatory vs. impairing hyperactive/impulsive behaviors that drove the current study's development (e.g., Kofler et al., 2016; Rapport et al., 2009).

Based on the DSM clinical model, each hyperactivity/impulsivity symptom is expected to confer impairment and be considered problematic by parents/teachers because it is in some way distracting/disruptive (Landau & Moore, 1991). To that end, a large body of evidence points to the importance of sensory modality in the extent to which external information is actively disruptive (Bettencourt & Xu, 2016). Interestingly, excess physical movement symptoms appear to be primarily visually-perceived (seen) by others (e.g., fidgets, leaves seat, runs about, etc.), whereas several other DSM hyperactivity/impulsivity items are primarily auditorily-perceived (heard) by others (e.g., talks excessively, blurts out). In that context, it is important to consider not only the distinct neural networks involved in the child's motor movement versus verbal speech production (Desmurgent & Sirigu, 2009; Soros et al., 2006; Tanji, 2001; van de Ven et al., 2009), but also how these behaviors are experienced by teachers, parents, and other children – the *informant's* perceptual modality. While visually-perceived movement in the environment can produce a rapid, automatic visual orienting response (Theeuwes et al., 1999), visual information is processed in anatomically distinct cortical regions from auditory/verbal information (Nee et al., 2013) and as such minimally disrupts verbal processing (Napolitano & Sloutsky, 2004; Rees et al., 2001). In contrast, even relatively brief verbal intrusions have the potential to significantly disrupt other children (as well as parents and teachers) because these intrusions directly compete for others' verbal processing resources. Indeed, research indicates that verbal intrusions interrupt children's selective attention (Wetzel & Schroger, 2014) and impede other students' performance on academic tasks (Dockrell & Shield, 2006). This impairment occurs because verbal intrusions gain automatic access to other children's phonological loops where they necessarily compete with verbal/auditory information currently held in mind (Baddeley, 2007).

Applying these perceptual models of sensory-specific interference effects on others' attention and memory, we hypothesized that excess physical movement may be less disruptive than verbal intrusions *to others* in most situations given developmental evidence that individuals in Western cultures show a strong preference for verbal processing (i.e., thinking and reasoning using language; e.g., Hitch et al., 1988). A key implication of this hypothesis is that conflating excess motor movement with verbally-loaded impulsivity/intrusion under a general 'hyperactivity/impulsivity' umbrella may be driving the negative associations between hyperactivity and functional outcomes in ADHD. As an example, it is easier to read silently when another child is quietly walking around the classroom than when that child is talking to their peer. Similarly, adults are better able to remain focused on a phone call (verbal) while doodling (visual) than while trying to read a manuscript (verbal).

However, directly testing our hypothesis that hyperactive/impulsive subclusters differentially predict functional outcomes would be premature because to our knowledge no study to date has examined whether it is psychometrically defensible to isolate excess physical movement (visual hyperactivity) from the verbally intrusive behaviors assessed under the ADHD 'hyperactivity/impulsivity' umbrella. A partial exception to this critique is a recent study by Gibbins et al. (2012), who modified their preplanned model to address unexpected factor loadings in an adult sample, and found the best fitting bifactor model distinguished between what they called motor hyperactivity/impulsivity and verbal hyperactivity/impulsivity. As shown below, this modified factor structure matches our rationally-derived structure with one exception. To our knowledge, however, this structure has not been tested in a child sample, and no study to date has taken a cognitively-informed, theoretically-driven approach to differentiate excess physical movement from other ADHD 'hyperactive/impulsive' behaviors or test the convergent/divergent validity of this alternative conceptualization.

Current Study

The current study addresses this omission and uses an empirically-informed rational approach (Clark & Watson, 1995) followed by confirmatory modeling and convergent/divergent validity testing. We hypothesized that (1) knowledgeable judges could reliably reclassify the current DSM hyperactivity/impulsivity item pool into excess physical movement (visual hyperactivity) and excess vocalizations or other noise that others perceive auditorily (verbal intrusions) subdomains, and that these rationally-derived subdomains would show adequate internal consistency reliability; (2) this alternative factor structure would provide reasonable model fit, latent reliability, and replicability based on both parent and teacher report; (3) this alternative bifactor structure would show convergent and divergent validity evidence, such that only visually-perceived hyperactivity would correlate significantly with objectively-measured excess physical movement (actigraphs); and (4) results would be robust to inclusion/exclusion of ADHD-inattentive symptoms.

Method

Open Data and Open Science Disclosure Statement

The de-identified dataset (.jasp), annotated results output (including test statistics), and lavaan analysis scripts are available for peer review: https://osf.io/qaxc2/. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012). The current study uses the same sample described in Kofler, Irwin et al. (2019). Actigraph data for a subset of the current sample was reported in aggregate to examine conceptually unrelated hypotheses in Kofler, Spiegel et al. (2018). None of the current study's outcome measures (item-level ADHD symptom data and raw actigraph scores) have been reported previously.

Participants

The sample included 132 clinically-evaluated children aged 8–13 years (M=10.34, SD=1.51; 47 girls) from the Southeastern U.S. recruited through community resources from 2015–2018 for participation in a larger study of the neurocognitive mechanisms underlying pediatric attention and behavioral problems (Table 1). Psychoeducational evaluations were provided to all caregivers. IRB approval was obtained/maintained, and all parents and children gave informed consent/assent. Sample ethnicity was mixed with 88 White/Non-

Hispanic (66.7%), 17 Hispanic/English-speaking (12.9%), 16 African-American (12.1%), 3 Asian (2.3%), and 8 multiracial children (6.1%).

All children and caregivers completed an identical, comprehensive psychoeducational and diagnostic evaluation that included a detailed, semi-structured clinical interview using the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS; Kaufman et al., 1997). The K-SADS (2013 Update) allows differential diagnosis according to symptom onset, course, duration, quantity, severity, and impairment in children and adolescents based on DSM-5 criteria (APA, 2013), and was supplemented with parent and teacher ratings from the Behavior Assessment System for Children (BASC-2/3; Reynolds & Kamphaus, 2015) and ADHD Rating Scale-4/5 (ADHD-4/5; DuPaul et al., 2016). ADHD diagnosis was conferred by the directing clinical psychologist based on (1) K-SADS; (2) borderline/clinical elevations on at least one parent and one teacher ADHD subscale; and (3) current impairment based on parent report. All ADHD subtypes/presentations were eligible given the instability of ADHD subtypes (Valo & Tannock, 2010). Comorbidities were included to improve generalizability and reflect clinical consensus best estimates (Kosten & Rounsaville, 1992). Psychostimulants ($n_{prescribed}$=25) were withheld ≥24 hours for testing; the washout duration was set based on the larger NIMH-funded study's approved protocol developed in consultation with the study's consulting psychiatrist. A psychoeducational report was provided to parents. Please see the larger study's preregistration for a detailed account of the comprehensive psychoeducational evaluation and study procedures (https://osf.io/2hmqp/).

The final sample included 132 children: 43 children with ADHD; 39 children with ADHD and common comorbidities (20 anxiety, 5 depression, 6 autism spectrum disorder/ASD, 8 oppositional-defiant disorder/ODD); 26 with common clinical diagnoses but not ADHD (15 anxiety, 3 depression, 7 ASD, 2 ODD); and 24 neurotypical children. The ADHD (n=82) and Non-ADHD samples (n=50) did not differ significantly in the proportion of children diagnosed with a clinical disorder other than ADHD (overall: p=.33, anxiety: p=.27, depression: p=.98, ODD: p=.23, ASD: p=.12). A subset of the participants screened positive for learning disorders in reading (9%) or math (7%); all of these participants also had diagnoses of ADHD. Children were excluded for gross neurological, sensory, or motor impairment; history of seizure disorder, psychosis, or intellectual disability; or non-stimulant medications that could not be withheld for testing.

Parent and Teacher-rated Hyperactivity

The ADHD Rating Scale for DSM-4/5 (ADHD-RS-4/5; DuPaul et al., 2016) forms each include the 18 DSM ADHD symptoms assessed on a 4-point scale (*never/rarely*, *sometimes*, *often*, *very often*). Higher scores reflect greater quantity/frequency of ADHD symptoms. Internal consistency in the current sample was ω=.94-.95 (teacher and parent) and parent/teacher interrater agreement was r=.45 (p<.001). Item-level responses for the 9 hyperactivity/impulsivity items were used for the study's primary analyses; the 9 inattentive items were used in exploratory analyses to probe the robustness of the final best fitting models. Parent and teacher ratings were modeled separately (Table 1).

Objectively-measured Hyperactivity

Actigraphs—Basic Motionlogger® (Ambulatory Monitoring, Inc., 2014) actigraphs are acceleration-sensitive devices that sample movement intensity 16 times per second (16 Hz), collapsed into 1-second epochs. The estimated reliability for actigraphs placed at the same site on the same person ranges from .90 to .99 (Tryon et al., 1991). Children were told that the actigraphs were "special watches" that let them to play the computerized learning games. Observer XT (Noldus, 2014) software was used to code start and stop times for each task, which were matched to the time stamps from the actigraphs. Children wore actigraphs on their non-dominant wrist and both ankles (i.e., 3 actigraph scores per child per task). Higher scores indicate greater intensity of movement (proportional integrating measure/PIM mode). A priori, we selected actigraph data collected during the Rapport et al. (2009) phonological and visuospatial working memory tasks, as well as during the beginning and end of session painting tasks. These tasks were administered as part of a larger battery of laboratory-based tasks that involved 1–2 sessions of approximately 3 hours each. The tasks were selected to provide a broad sampling of children's gross motor movement across both high and low cognitive demands and given precedence for using actigraphy to measure children's activity level during these tasks (e.g., Alderson et al., 2012; Kofler et al., 2018; Rapport et al., 2009). All tasks were counterbalanced to minimize order effects. Children received brief breaks after each task, and preset longer breaks every 2–3 tasks to minimize fatigue. Detailed working memory task descriptions and performance data for this sample are reported in Kofler, Irwin et al. (2019). The paint condition involved children using Microsoft Paint for five consecutive minutes at the beginning and end of the first research session. Children sat in the same caster-wheel swivel chair and interacted with the same computer used for the working memory tasks while using a program that placed relatively modest demands on executive processes (i.e., Paint allows children to draw/paint on the monitor using a variety of interactive tools). Performance was monitored at all times by the examiner, who was stationed just outside of the testing room (out of the child's view) to provide a structured setting while minimizing performance improvements associated with examiner demand characteristics (Gomez & Sanson, 1994).

Bifactor-(s-1) Models

The bifactor model was selected *a priori* given our goal of maximally differentiating excess physical movement (visual hyperactivity) from other indicators of ADHD-related hyperactivity/impulsivity. Following recommendations for bifactor models by Eid et al. (2018b), the current study used a bifactor-(s-1) structure such that each indicator loaded onto a general factor (i.e., hyperactivity) and a subset of indicators also loaded onto a specific factor (e.g., verbal intrusion). As required to properly fit the bifactor model and interpret the general factor, one or more items must load onto the general factor but not onto any specific factor (Eid et al., 2018a). These reference facets serve as markers that define the meaning of the general factor (in this case, hyperactivity). In the current study, the general factor was reified 'visual hyperactivity' because its reference facets specifically assess excess physical movement (fidgets, leaves seat, runs/climbs, always on the go). Visual hyperactivity was selected *a priori* as the general factor given our primary interest in excess physical movement, the limited number of DSM 'impulsivity' items, and the historical emphasis on excess physical movement as central to diagnostic and conceptual models of ADHD (reviewed above); alternative factor structures were also tested and are reported below and in the supplementary materials.

Importantly for our purposes, the general factor is modeled as uncorrelated with the specific factor(s) in the bifactor model based on the underlying assumption that an individual's score on an item reflects at least two distinct sources of reliable variance (i.e., attributable to the general factor and the specific factor). That is, the bifactor model was applied based on our *a priori* hypotheses that (a) informant ratings on a subset of DSM-5 hyperactivity/impulsivity items reflect reliable variance attributable primarily to excess physical movement (i.e., fidgets, leaves seat, runs/climbs, always on the go), whereas the remaining items may be influenced by children's physical movement but also contain reliable variance attributable to impulsivity or verbal intrusion; and (b) these perceptions can be parsed into separate, latent estimates, with the resultant visual hyperactivity factor providing a reliable, construct-level indicator of children's excess physical movement that would show convergent validity with objectively-assessed physical activity (Kofler et al., 2016).

Data Analysis Overview

An empirically-driven rational approach was used to develop the models (Clark & Watson, 1995), which were then tested using confirmatory factor analysis (CFA). Our primary and exploratory analyses are organized into four Tiers that involved expert judges' classification of each DSM hyperactivity/impulsivity item (Tier 1), confirmatory factor analyses comparing the judges' rationally-derived factor structure to extant models of hyperactivity/impulsivity (Tier 2), tests of the convergent and divergent validity of Tier 2's best-fitting model (Tier 3), and sensitivity analyses to probe the impact of key methodological decisions on study results.

For all confirmatory models, absolute and relative fit were tested. Adequate model fit is indicated by CFI and TFI ≥ .90, and RMSEA ≤ .10. AIC and BIC were used to compare non-nested models; smaller values indicate the preferred model (Kline, 2016). BIC reductions of 0–6 are considered positive support for the preferred model (6–10=strong support, >10=decisive support; Kass & Raftery, 1995). Omega total (ω) and omega subscale (ωs) index the reliability of the general factor (visual hyperactivity) and specific factor (impulsivity or verbal intrusion) by providing estimates of the proportion of observed score variance attributable to sources of common and specific variance, respectively; values >.70 are preferred (Rodriguez et al., 2016b). Omega hierarchical (ω_H) and omega subscale hierarchical (ω_{HS}) estimate the proportion of reliable variance in observed scores attributable to the general factor after accounting for the specific factor, and to the specific factor after accounting for the general factor, respectively. Explained common variance (ECV) indicates the proportion of reliable variance explained by each factor. The percentage of uncontaminated correlations (PUC) is used to assess potential bias from forcing multidimensional data into a unidimensional model. When general factor ECV > .70 and PUC > .70, bias is considered low and the instrument can be interpreted as primarily unidimensional (i.e., the increased complexity of the bifactor structure is likely not warranted; Rodriguez et al., 2016a). Construct replicability (H) values > .80 suggest a well-defined latent variable that is more likely to be stable across studies (Watkins, 2017).

All items were screened for normality (all skewness and kurtosis < 1.5) and showed the full range of response options (range=0–3). Standardized residuals were inspected for magnitude (all positive and < 1, indicating no evidence of localized ill fit). Directionality of parameter estimates were inspected. Completely standardized theta scaling parameterization (i.e., delta scaling) with maximum likelihood estimation with robust standard errors (MLR) was used to account for the ordinality of the data (Kline, 2016). MLR is considered appropriate for ordered categorical data with greater than three response options (Green et al., 1997) and was selected because it provides the critical AIC and BIC tests needed to compare our non-nested models.

Results

Power Analysis

A series of Monte Carlo simulations were run using Mplus7 (Muthén & Muthén, 2012) to estimate the power of our proposed bifactor model for detecting significant factor loadings of the expected magnitude, given power $(1-\beta) \geq .80$, $\alpha=.05$, and 10,000 simulations per model run. Briefly, this process compiles the percentage of model runs that result in statistically significant estimates of model parameters. Standardized factor loadings and expected residual variances for observed variables were imputed iteratively to delineate the proposed bifactor model. Results indicated that our sample size of 132 is powered to detect standardized factor loadings ≥ .46, which falls at/below the standardized factor loadings reported in most ADHD bifactor models (e.g., Allan & Lonigan, 2018; Quyen et al., 2017; Ullebø et al., 2012). Thus, the study is sufficiently powered to address our primary aims.

Tier 1. Empirically-Driven Rational Approach to Model Development

Given our goal of assessing the feasibility and utility of maximally differentiating excess physical movement (visual hyperactivity) from other indicators of ADHD-related hyperactivity/impulsivity, we first examined the current DSM item pool to determine if there were a sufficient number of items falling into each subdomain. To this end, the 9 DSM hyperactivity/impulsivity items were judged to fall into one of three categories using an empirically-driven rational approach (Clark & Watson, 1995). The seven judges independently determined whether each item reflected (1) excess physical movement that others observe visually (reified ‘visual hyperactivity’), (2) excess vocalizations or other noise that others perceive auditorily (reified ‘verbal intrusion’), or (3) both. A fourth category, ‘neither’, was dropped from analyses because it was unused by all judges.

Intraclass correlation and Fleiss’ kappa were computed to test the reliability of the revised scale categories (Nunnally & Bernstein, 1994) using the R functions ICC and fleissm.kappa, respectively (from packages psych and irr; Gamer et al., 2019; Revelle, 2018). The intraclass correlation was computed using Shrout and Fleiss (1979; Case 2) to determine the generalizability of the judges’ item categorizations and indicated excellent agreement: ICC(8,48)=.95, 95%CI=.88-.99, $p<.0005$. Fleiss’ kappa for more than two raters (Siegel &

Castellan, 1988) provided further evidence of reliability, κ=.85, p<.0005. Internal consistency for the rationally-derived visual hyperactivity/verbal intrusion subdomains was excellent in the current sample based on both parent (ω=.87-.89, α=.87-.89) and teacher report (ω=.88-.95, α=.84-.95). Items judged to belong to each category are shown in Table 2.

Tier 2a: Informant Measurement Models

In Tier 2a, we used confirmatory factor analysis (CFA) via the R package lavaan (JASP Team, 2018; Rosseel, 2012) and the software program Omega (Watkins, 2017). Five potential models were tested for each informant: (1) the traditional DSM-5 single-factor model (all 9 hyperactivity/impulsivity items loading onto one general factor); (2) the traditional DSM-IV 2-factor model, with correlated hyperactivity (items 1–6) and impulsivity factors (items 7–9); (3) the DSM-IV conceptualization as a bifactor model (items 7–9 forming a unique impulsivity factor); (4) our alternative conceptualization as a 2-factor model, with correlated visual hyperactivity (items 1–3 and 5) and verbal intrusion factors (items 4 and 6–9); and (5) our visual hyperactivity/verbal intrusion conceptualization as a bifactor model (described above). Parent and teacher reports were modeled separately.

As shown in Tables 3–4, the DSM-5 single-factor model showed inferior fit relative to all 2-factor correlated and bifactor models based on both parent and teacher reports (ΔBIC = 20–44, ΔAIC = 28–58; values > 10 indicate decisive support; Kass & Raferty, 1995). Similar to prior studies, however, the 2-factor correlated DSM-IV hyperactivity/impulsivity model (r=.81-.93, p < .001) and the alternative 2-factor correlated visual hyperactivity/verbal intrusion model (r=.83-.90, p < .001) both produced factors that were highly correlated/ multicollinear. This is a common finding in prior work, where the correlated factors model fits better than the single-factor model but is ultimately rejected in favor of parsimony due to the high correlation between factors (for review, see Allan & Lonigan, 2019).

Inspection of the best-fitting bifactor models indicated that this multicollinearity is likely attributable, at least in part, to forcing items that are influenced by both constructs to be reflective of only one of those constructs. That is, the bifactor approach creates factors that are uncorrelated by design, allowing theoretically-specified item subsets to contribute to both factors. In that context, the best fitting model was the visual hyperactivity/verbal intrusion bifactor model (Figure 1; Tables 3–4). This finding replicated across the parent- and teacher-report data. The evidence for this model's superior fit was decisive relative to the DSM-5 single factor model (ΔBIC=26–44; ΔAIC=40–58), positive to decisive relative to the DSM-IV bifactor hyperactivity/impulsivity model (ΔBIC=8–11; ΔAIC=2–6), positive to decisive relative to the DSM-IV correlated factors hyperactivity/impulsivity model (ΔBIC=1–12; ΔAIC=5–12), and positive to decisive relative to the visual hyperactivity/ verbal intrusion correlated factors model (ΔBIC=1–2, both ΔAIC=12).

Inspection of the factor reliability and replicability indices for the preferred visual hyperactivity/ verbal intrusion bifactor models indicated that the percent of uncontaminated correlations was <.70 in both the parent and teacher models (both PUC=.44), supporting the multidimensionality of the item set. In addition, there was support for the reliability of the general visual hyperactivity (ω=.94-.96) and specific verbal intrusion factors (ωs=.88-.95) in both parent and teacher models. Total variance explained by the models was .54 (parent) to .63 (teacher). Explained common variance and factor-specific variance explained was high for the general hyperactivity factor (ECV=.84-.88, ω_H=.86-.90), but relatively low for the specific verbal intrusion factor in both parent and teacher models (ECV=.12-.16, ω_{HS}=.17-.26), indicating that the specific factor explained a meaningful but modest proportion of variance relative to the general factor. Construct replicability was high for the general hyperactivity factor (H=.92-.95) but was below target thresholds for the specific factor (H=.48-.56).[1]

Taken together, the visual hyperactivity/verbal intrusion bifactor models showed superior fit, adequate to excellent reliability, and evidence supporting multidimensionality based on both parent and teacher data. The visual hyperactivity factor explained a large proportion of common variance and showed excellent construct replicability. In contrast, the verbal intrusion factor explained a modest but likely meaningful proportion of common variance (Canivez, 2015), but lower replicability, suggesting that the extant item set is multidimensional but likely provides insufficient coverage of the verbal intrusion construct for interpretation of its factor score in applied settings (Watkins, 2017).

Given the uniformly strong support for the visual hyperactivity factor that was of primary interest in the current study, the visual hyperactivity/verbal intrusion bifactor model was retained in Tiers 3–4, with the caveat that interpretation of relations with the verbal intrusion factor should be interpreted with caution due to questionable reproducibility based on the available item set.

Tier 2b: Actigraph Measurement Models

The actigraph bifactor model was also built in Tier 2 in preparation for its use in Tiers 3–4 as a test of the convergent/divergent validity of the informant-report models described above. Two models were tested: a single-factor model (all actigraph data points loading onto a general hyperactivity factor), and a bifactor model that included baseline hyperactivity (general factor reflecting activity level common across tasks) and task-specific hyperactivity (specific factors reflecting activity level related specifically to each task). As shown in Table 5, the single-factor actigraph model provided poor fit to the data. In contrast, the bifactor actigraph model that included both baseline hyperactivity (general factor) and task-specific hyperactivity (specific factors) provided adequate model fit (Figure 2; Table 5). This model was consistent with meta-analytic findings (Kofler et al., 2016) indicating that children exhibit a baseline (general) level of physical/motor movement that is modulated by task-specific demands in the environment (Figure 2).

Inspection of the factor reliability and replicability indices for the actigraph bifactor model indicated that the percent of uncontaminated correlations was high (PUC=.80) but the general factor's explained common variance was low (ECV=.42), supporting the multidimensionality of the data. In addition, there was support for the reliability of the

[1]We also ran these analyses for the bifactor hyperactivity/impulsivity parent and teacher models. Values were generally lower than those reported in the main text, providing further support for selecting the visual hyperactivity/verbal intrusion bifactor models; detailed results are posted on our OSF website [linked above].

general hyperactivity (ω=.97) and all specific factors (ωs=.93-.94). Explained common variance was moderate for the general factor (ECV=.42) and the specific factors (ECV=.18-.21), suggesting meaningful contributions of all modeled factors. Finally, construct replicability was above thresholds for the general factor (H=.97) and all specific factors (H=.83-.91). The bifactor actigraph model was therefore retained in Tiers 3–4.

Tier 3: Structural Models

In Tier 3, we tested the hypothesis that objectively-assessed hyperactivity would show stronger associations with informant-rated hyperactivity than with informant-rated verbal intrusion, to the extent that our alternative conceptualization of these symptom clusters has ‘real world’ applicability. To accomplish this goal, the Tier 2b actigraph factors were correlated with the Tier 2a visual hyperactivity and verbal intrusion factors (separately for parents and teachers). We then used tests of dependent correlations as implemented in the R package cocor (Diedenhofen & Musch, 2015) to test whether the association with actigraph-measured visual hyperactivity was significantly higher for informant-reported visual hyperactivity than for informant-reported verbal intrusion as hypothesized.

The models showed adequate to excellent fit (Tables 6–7). Results were highly consistent across the parent and teacher models, such that the informant-rated visual hyperactivity factor was significantly associated with the objective, actigraph-measured baseline hyperactivity factor (r=.25-.27, p<.005) as predicted. Interestingly, the informant-rated visual hyperactivity factor was also significantly associated with task-specific hyperactivity during the visuospatial working memory task (r=.19-.25, p<.03) but not during the phonological working memory (r= −.08 to .08, both p>.35) or computer painting activities (r=.05-.07, p>.30). The models’ divergent validity was also supported: Informant-rated verbal intrusion was not associated with objectively-measured hyperactivity either overall (r= −.03 to .08, p>.44) or during any specific task (r= −.13 to .15, p>.17).

Taken together, the results provided both convergent and divergent validity evidence, such that objectively-measured hyperactivity was significantly and uniquely associated with parent- and teacher-reported visual hyperactivity only. However, it is not appropriate to conclude that one correlation is significantly larger than another based on a difference in significance level alone. We therefore used tests of dependent correlations (Diedenhofen & Musch, 2015) to test whether the associations with actigraph-measured baseline (general) hyperactivity was significantly higher for informant-reported visual hyperactivity than for informant-reported verbal intrusion as hypothesized. Results indicated that the difference in correlation magnitude was significant in the parent model (r= .25 vs. −.03, p = .009), whereas this difference failed to reach significance for the teacher model (r= .27 vs. .08, p = .055).

Tier 4: Sensitivity Analyses

In Tier 4, we probed the impact of key methodological decision points on study results. Results are reported in detail in the Supplementary Online materials and summarized here.

First, we probed the extent to which results were impacted by our *a priori* decision to focus on the DSM hyperactivity/impulsivity items (i.e., to exclude the DSM inattention items).

This involved adding the 9 DSM inattention items to the best fitting Tier 2 parent and teacher models. We tested each model twice: once with the inattention items forming a separate factor correlated with the visual hyperactivity/verbal intrusion bifactors, and once with the inattention items added to the bifactor model (with visual hyperactivity retained as the general factor given our goal of maximally differentiating it from verbal intrusion, and the addition of a specific inattention factor indicated by DSM inattention items 1–9).[2] As shown in Supplementary Table S1, the pattern and interpretation of results is unchanged with inattentive symptoms added to the model, with the possible exception of stronger support for the visual hyperactivity/verbal intrusion bifactor structure's convergent/divergent validity based on teacher report (relative to uniformly strong support across parent models).

Next, we tested the convergent/divergent validity of the DSM-5 single-factor and DSM-IV hyperactivity/impulsivity bifactor models relative to actigraph-measured hyperactivity. Although these models showed inferior fit to the preferred visual hyperactivity/verbal intrusion bifactor model in Tier 2, the construct coverage and replicability of the verbal intrusion factor was questionable even in the preferred model. Therefore, these exploratory models were run to probe the extent to which the preferred model provided improved assessment of excess physical movement relative to extant models as hypothesized. As shown in Supplementary Table S2, the pattern and interpretation of results was similar to the preferred Tier 3 findings, with the exception that both alternative options diverged in theoretically unexpected ways. In particular, DSM-5 hyperactivity/impulsivity was not significantly associated with actigraph-measured hyperactivity in the parent model, suggesting limited use for assessing excess motor movement as intended. Further, the DSM-IV impulsivity factor showed atheoretical associations with actigraph-measured hyperactivity, such that higher impulsivity was associated with both lower and higher actigraph-measured movement across conditions. In addition, inspection of model fit, reliability, and replicability indices indicated that these exploratory structural models showed positive to decisive evidence for inferior fit relative to the preferred Tier 3 structural model.

Finally, we probed our *a priori* decision to model visual hyperactivity as the general factor. This involved re-fitting the visual hyperactivity/verbal intrusion bifactor models, this time with verbal intrusion modeled as the general factor and visual hyperactivity as the specific factor. As described in the Supplementary materials, these models showed similar model fit, reliability, and reproducibility estimates relative to the preferred/best-fitting models, but produced anomalous convergent/divergent validity results that question their interpretability.

[2]This latter model is the most similar to prior bifactor models of the 18 DSM-5 ADHD items, with the exception that we retained hyperactivity as the general factor given our primary goal of maximally differentiating excess physical movement from other ADHD symptoms. Of note, prior ADHD studies that have used bifactor modeling have interpreted the general factor as 'ADHD.' However, as shown in Eid et al. (2018a, 2018b), to our knowledge all prior ADHD bifactor models have been misspecified, thus obscuring interpretation of the general factor due to the lack of reference facets to define its meaning. In the current study, the meaning of the general factor is defined as 'visual hyperactivity' because its reference facets clearly assess excess physical movement (i.e., fidgets, leaves seat, runs/climbs, always on the go). We considered testing a model with a general 'ADHD' factor and specific inattention, hyperactivity, and impulsivity/verbal intrusion factors. However, we made the *a priori* decision not to do so because (1) questions regarding the factor structure of ADHD as a diagnostic entity are beyond the study's scope; (2) the results already indicated insufficient item coverage for key constructs (i.e., verbal intrusion/impulsivity); and (3) we had concerns regarding power and localized ill fit for the complex model that includes a general and 3 specific factors due to (a) the low indicator:factor ratio, and (b) insufficient item counts for at least some models secondary to the need to have at least one reference facet from each domain not load on any specific factor to properly define the general factor's meaning as 'ADHD' (e.g., there would be only 2 items on the 'impulsivity' factor, which is problematic for model construction; Kline, 2016).

Consistent with our theoretical assumptions described above, these unexpected findings suggest that the general factor continues to primarily assess visual hyperactivity despite our attempt to define it as verbal intrusion, but that it does so less precisely relative to the preferred fitting model described above.

Discussion

The current study was the first to combine rational and empirical approaches to apply cognitive-perceptual models to the study of overt behavior and maximally differentiate excess physical movement (hyperactivity) from other behaviors assessed under the ‘hyperactivity/impulsivity’ umbrella. Additional strengths include the relatively large, clinically evaluated sample, multi-informant replication, and inclusion of objective assessment of hyperactivity (actigraphy) that provided strong tests of convergent validity. We found that the current DSM item pool can be reliably reclassified by knowledgeable judges into items reflecting excess physical movement (visual hyperactivity) and auditory interruption (verbal intrusion) symptoms, and that these rationally-derived subdomains show excellent internal consistency reliability. This alternative bifactor structure showed evidence for multidimensionality as hypothesized and superior model fit relative to traditional hyperactivity/impulsivity models. The resultant excess physical movement (visual hyperactivity) factor was reliable, replicable, and showed strong convergent validity evidence via its association with objectively-assessed hyperactivity that replicated across informant models. The verbal intrusion factor also showed evidence for reliability and explained a substantive portion of reliable variance, but demonstrated insufficient construct coverage as evidenced by low estimated replicability.

Despite lower replicability, the verbal intrusion factor accounted for a reasonable portion of reliable variance in both parent and teacher models (12%−16%), and in doing so improved the specificity of the excess physical movement factor as evidenced by more empirically and theoretically consistent links with objectively-assessed physical movement. These findings collectively suggest that the DSM-5 hyperactivity/impulsivity item pool is multidimensional and includes high levels of reliable variance associated with informant perceptions of children’s excess physical movement. In addition, informant ratings are influenced to a significant extent by children’s verbally intrusive behavior. ‘Verbal intrusion’ appears to provide a better explanation than ‘impulsivity’ for the reliable, non-hyperactivity variance assessed by these items, but the current item set appears insufficient for replicable measurement of this construct. That is, the current DSM-5 item set does not appear to adequately assess verbal intrusion to a degree that justifies independent clinical interpretation, despite accounting for enough reliable variance to improve the construct validity of assessing excess physical movement via informant report. Taken together, the current findings are consistent with results from a recent adult study (Gibbins et al., 2012) and highlight the feasibility and potential utility of approaches that maximally distinguish visually-perceived excess physical movement from other behaviors captured under the hyperactivity/impulsivity umbrella. To that end, and in consideration of the need for psychometric development work to adequately capture the hypothesized verbal intrusion construct, our discussion below expands on the empirical basis for developing this line of inquiry and building testable hypotheses for future work to refine the nature, mechanisms underlying, and functional outcomes of excess physical movement in ADHD.

The current study differentiated excess physical movement from verbally-mediated intrusions/interruptions based on neurocognitive literature indicating that these behaviors are supported by distinct cortical networks and show differential effects *on others* as a function of their sensory and memory processes (Napolitano & Sloutsky, 2004; Nee et al., 2013; Rees et al., 2001). This approach has the potential to improve our understanding of ADHD-related symptoms and how these symptoms produce or mitigate the disorder's hallmark functional impairments (Willcutt et al., 2012). For example, this distinction may be particularly helpful for unpacking impairments in peer and family functioning by providing a mechanism by which the child with ADHD's behavior produces impairments for others (which in turn contributes to the interpersonal difficulties commonly associated with ADHD; Gaub & Carlson, 1995). Along these lines, we hypothesize that the higher levels of physical movement associated with ADHD may be less impairing to children with ADHD because this movement is simultaneously more likely to facilitate cognition for children with ADHD (Pontifex et al., 2015; Sarver et al., 2015) and less likely to interfere with parents', teachers', and other children's task-related verbal processing. This hypothesis is based on replicated evidence that physical movement is processed in anatomically distinct brain regions from verbal information (e.g., Nee et al., 2013).

In contrast, school-aged children in Western cultures show a strong preference for verbal processing (i.e., thinking and reasoning using language; e.g., Hitch et al., 1988). Thus, we further hypothesize that verbal intrusion symptoms may be maximally associated with interpersonal impairment to the extent that these behaviors – by their very nature – actively interfere with what other children are trying to do. For example, consider these behavioral symptoms as they are expressed in classroom settings. Replicated evidence from dual-dissociation studies indicates that new information from the environment (e.g., a child with ADHD talking out of turn) actively interferes with the encoding, storage, and processing of same-modality information necessary for goal-directed behavior (e.g., peers trying to read silently, teachers trying to lecture; Banbury et al., 2001; Postle et al., 2005). In contrast, cross-modality information (e.g., a child with ADHD quietly getting out of her seat) tends to show much lower interference effects (Berti & Schroger, 2001). In other words, it may be fairly easy for other children to ignore a child with ADHD during independent seatwork if she is quietly walking around the classroom, whereas it would be more difficult to tune her out if she were singing the passage she was supposed to be reading silently. Of course, these hypotheses were not tested in the current study and thus remain speculative and dependent on the outcome of future psychometric work aimed at developing methods for reliable and valid assessment of verbal intrusion.

Admittedly, we were somewhat surprised that our alternative conceptualization fit the data *better* than prevailing hyperactivity/impulsivity models, particularly given that the DSM item pool used in the current study was selected based on prior factor analyses (e.g., Lahey et al., 1994; Willcutt et al., 2012) that have been replicated in many studies (for review see Allan & Lonigan, 2018). When beginning this project, we conceptualized it as a 'proof of concept' first step, with the expectation that the findings would indicate the need for larger item sets to adequately separate visual hyperactivity from verbal intrusion. Indeed, this expectation was partially supported. Taken together with recent results from an adult sample (Gibbins et al., 2012), the current findings suggest that the current DSM-5 item pool may be sufficient for screening for ADHD-related symptoms of excess physical movement (called motor hyperactivity/impulsivity in Gibbins et al., 2012). In contrast, the most parsimonious conclusion appears to be that the verbal intrusion construct is present in the DSM-5 item set to a sufficient degree to confound measurement of visual hyperactivity but not to a sufficient degree that it can be independently interpreted in applied/clinical settings at this time.

Limitations

Several caveats merit consideration despite our study's rational-empirical approach, use of multi-informant measures, and evaluation of convergent and divergent validity via objective assessment of hyperactivity. Most notably, we were unable to include an objective index of verbal intrusion, which would have been a larger limitation had the DSM-5 item pool sufficiently captured this construct. The utility of our proposed reconceptualization would be strengthened significantly by future work testing for dual dissociation between subjective and objective measures of visual hyperactivity and verbal intrusion. As such, psychometric work is needed to develop both subjective and objective measures of verbal intrusion (e.g., decibel level, adapting behavioral codes such as vocalizations; Platzman et al., 1992), toward specifying its construct space and relations with functional impairments in ADHD.

Further, our characterization of excess physical movement as 'compensatory' is based on replicated empirical findings (Hartanto et al., 2015; Pontifex et al., 2015; Sarver et al., 2015; Rapport et al., 2009) but is likely an oversimplification of a more complex clinical picture. Indeed, it is likely that excess physical movement can be both compensatory (e.g., movement while remaining engaged in a task, such as a child bouncing in her chair while reading) and impairing (e.g., movement that takes the child away from the task at hand). Similarly, it is likely that some behaviors characterized as intrusive or impulsive may be compensatory (e.g., blurting out or interrupting to convey an idea to others before it is forgotten; using external private speech to compensate for the child's underdeveloped working memory system; Dickman, 1990; Rapport et al., 2001). Further, despite neurocognitive evidence that verbally-mediated information processing is more strongly disrupted by auditorily-perceived information than by visually-perceived information, it is likely that visually-perceived behaviors may be more disruptive in other circumstances (e.g., in environments where physical safety is a concern). Expanding the model to include compensatory and intrusive/impulsive categories, each with physical movement and auditory subscales, was beyond the scope of the study given the limited number of DSM hyperactivity/impulsivity items but will be important for maximally targeting impairing behaviors while facilitating compensatory actions.

Consistent with findings from our 2-factor correlated models, several previous studies have shown improved model fit when separating hyperactivity and impulsivity but nonetheless adopted the combined hyperactivity/impulsivity model based on the rule of parsimony (e.g., Allan & Lonigan, 2019; Burns et al., 2001; Toplak et al., 2009). Such a conclusion was considered here but rejected given that (1) the bifactor model produced uncorrelated factors with superior model fit relative to the single-factor and correlated 2-factor models; and (2) the increased model complexity resulted in not only improved model fit but importantly improved prediction to objective indicators. It will be important for future studies of ADHD's factor structure to include objective convergent/divergent validity outcomes to weigh the principle of parsimony against the potential for differential outcome prediction. Similarly, our proposed re-conceptualization of these symptoms based on informant perceptual processes and neurocognitive effects is based on a strong evidence base regarding differential cortical activation and interference as a function of sensory input modality. However, the current study did not test the extent to which the proposed visual hyperactivity and verbally intrusive behaviors in ADHD specifically resulted in differential cortical activation in observers/informants. This reflects an important area for future research.

Finally, our study was limited by the use of DSM symptom checklists, which provided a circumscribed number of items that failed to capture the full range of at least some of the constructs of interest. This limitation is shared by all modern factor analytic studies of DSM-based ADHD informant ratings, and highlights the clear need for objective symptom assessment (e.g., actigraphy) and prospective psychometric development studies to provide more complete surveys of ADHD behavioral subdomains (Burns et al., 2001; Lahey et al., 1998). In this context, it might be tempting to minimize the current study's empirical findings. That is, despite evidence for superior fit and improved specificity for assessing excess physical movement, there was insufficient coverage of the verbal intrusion construct and in practical terms the preferred model and the DSM-IV hyperactivity/impulsivity model differ by only a few items. Similarly, despite evidence that the actigraph data correlated significantly stronger with visual hyperactivity than verbal intrusion, the absolute magnitude of these associations was only moderate ($r \approx .3$), suggesting that processes beyond just how much a child moves are influencing parent and teacher perceptions that these children are moving excessively. Despite our method of using multiple actigraphs to more broadly capture motor movement, this level of convergence likely reflects the multitude of differences between objective, laboratory-based assessment and subjective, home- and school-based perceptions of motor movement (e.g., different settings and time frame; different susceptibility to recency, halo, retrospective recall, and other effects; and different potential impact of the observer's interpretations regarding the movement's intent and the extent to which it was disruptive and/or appropriate to the context). Our view is that the current study's contribution is primarily conceptual and foundational: Whether the term 'verbal intrusion,' 'verbal impulsivity,' or 'verbal hyperactivity' is preferred, the current findings and those of Gibbins et al. (2012) highlight the feasibility and potential utility of differentiating excess physical movement (visual hyperactivity) from other behaviors assessed under the 'hyperactivity/impulsivity' umbrella.

Clinical and Research Implications

As an initial study, the current findings provide strong preliminary support for the feasibility and potential utility of maximally differentiating informant perceptions of excess physical movement (visual hyperactivity) from other indicators of ADHD-related hyperactivity/impulsivity. Doing so may provide a useful heuristic for continuing to refine the nature, mechanisms underlying, and functional outcomes of excess physical movement in ADHD. If replicated and differentially linked with outcomes, this line of work may help clarify intervention targets (e.g., potential benefits of adopting recent ‘squirm to learn’ or ‘freedom to move’ recommendations that allow higher levels of physical movement while still targeting verbal intrusions that are disruptive to peers; Sarver et al., 2015). In addition, reconceptualizing these symptoms may have important implications for diagnostic decision-making and understanding ADHD-related impairment – particularly given replicated evidence that visual information minimally disrupts task-related verbal processing (e.g., Nee et al., 2013). This impairment for others may in turn predict the functional impairment (e.g., peer and family conflict) that drives clinical referrals for many children with ADHD (Pelham et al., 2005). In that context, the informant’s own level of frustration tolerance, multitasking abilities, executive functioning, etc. may be important considerations when integrating their ratings into the diagnostic decision-making process. That is, the quantity, frequency, and/or severity of a child’s reported ADHD symptoms may be related, at least in part, to individual differences in the informant’s experience of these behaviors as distracting/disruptive (De Los Reyes & Kazdin, 2005). These conclusions are consistent with preliminary evidence that caregivers with elevated depressive symptoms tend to inflate the quantity and severity of their child’s ADHD symptoms (Carrington et al., 2020), but must be considered speculative because the current study did not examine diagnostic decision-making. Nonetheless, these findings highlight the importance of considering aspects of the informant beyond their symptom ratings, as well as considering the potential utility of objective observational and/or mechanical data (e.g., actigraphy) for improving the science and technology of ADHD assessment in clinical practice (Rapport et al., 2000). Of course, these hypotheses remain speculative, but appear promising given the positive preliminary evidence identified herein.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in part by NIH grants (R34 MH102499-01, R01 MH115048; PI: Kofler). The sponsor had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

References

Abrams R, & Christ S (2003). Motion onset captures attention. Psychological Science, 14, 427–432.

Aduen PA, Day TN, Kofler MJ, Harmon SL, Wells EL, & Sarver DE (2018). Social problems in ADHD: Is it a skills acquisition or performance problem?. Journal of Psychopathology and Behavioral Assessment, 40, 440–451.

Ahmed GM, & Mohamed S (2011). Effect of regular aerobic exercise on behavioral, cognitive and psychological response in patients with attention deficit-hyperactivity disorder. Life Science Journal, 8(2), 366–371.

Alderson RM, Rapport MD, Kasper LJ, Sarver DE, & Kofler MJ (2012). Hyperactivity in boys with ADHD: The association between deficient behavioral inhibition, attentional processes, and objectively measured activity. Child Neuropsychology, 18, 487–505.

Allan DM, & Lonigan CJ (2018). Examination of the structure and measurement of inattentive, hyperactive, and impulsive behaviors from preschool to grade 4. Journal of Abnormal Child Psychology, 1–13.

Baddeley A (2007). Working memory, thought, and action. Oxford: OUP.

Banbury SP, Macken WJ, Tremblay S, & Jones DM (2001). Auditory distraction and short-term memory: Phenomena and practical implications. Human Factors, 43, 12–29.

Berti S & Schroger E (2001). A comparison of auditory and visual distraction effects: Behavioral and event-related indices. Cognitive Brain Research, 10, 265–273.

Bettencourt CB & Xu Y (2016). Decoding the content of visual short-term memory under distraction in occiptital and parietal areas. Nature Neuroscience, 19, 150–157.

Brabham EG & Lynch-Brown C (2002). Effects of teachers' reading-aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. Journal of Educational Psychology, 94, 465–473.

Brocki KC, Tillman CM, & Bohlin G (2010). CPT performance, motor activity, and continuous relations to ADHD symptom domains. European J Developmental Psychology, 7, 178–197.

Burns GL, Boe B, Walsh JA, Sommers-Flanagan R, & Teegarden LA (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.

Bunford N, Brandt EN, Golden C, Dykstra JB, & Suhr JA (2015). ADHD symptoms mediate the association between deficits in executive functioning and social impairment in children. Journal of Abnormal Child Psychology, 43, 133–147.

Chang Y, Liu S, Hou-Hsiang Y, & Yuan-Hung L, N.J. (2012). Effect of acute exercise on executive function in children with attention deficit hyperactivity disorder. Archives of Clinical Neuropsychology, 27, 225–237.

Clark LA & Watson D (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7(3), 309–319.

De Los Reyes A, & Kazdin AE (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. Psychological Bulletin, 131, 483–509.

Diedenhofen B & Much J (2015). cocor: A comprehensive solution for the statistical comparison of correlations. PLOS ONE, 10, 1–12.

Dockrell JE, & Shield BM (2006). Acoustical barriers in classrooms: The impact of noise on performance in the classroom. British Educational Research Journal, 32, 509–525.

DuPaul GJ, Power TJ, Anastopoulos AD, & Reid R (2016). ADHD rating scale-5 for children and adolescents: checklists, norms, and clinical interpretation. New York: Guilford.

Eid M, Geiser C, Koch T, & Heene M (2018). Anomalous results in G-factor models: Explanations and alternatives. Psychological Methods, 22, 541–562.

Eid M, Krumm S, Koch T, & Schulze J (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. J Intelligence, 6, 42.

Elliott JG, Gathercole SE, Alloway TP, Holmes J, & Kirkwood H (2010). An evaluation of a classroom-based intervention to help overcome working memory difficulties and improve long-term academic achievement. Journal of Cognitive Education and Psychology, 9, 227–250.

Gapin JI, Labban JD, & Etnier JL (2011). The effects of physical activity on attention deficit hyperactivity disorder symptoms: The evidence. Preventive Medicine, 52, 570–574.

Gaub M & Carlson C (1997). Behavioral characteristics of DSM-IV ADHD subtypes in a school-based population. Journal of Abnormal Child Psychology, 25, 103–111.

Gibbins C, Toplak ME, Flora DB, Weiss MD, & Tannock R (2012) Evidence for general factor model of ADHD in adults. Journal of Attention Disorders, 16(8), 635–644.

Green SB, Akey TM, Fleming KK, Hershberger SL, & Marquis JG (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. Structural Equation Modeling: A Multidisciplinary Journal, 4, 108–120.

Hartanto TA, Krafft CE, Iosif AM, & Schweitzer JB (2016). A trial-by-trial analysis reveals more intense physical activity is associated with better cognitive control performance in ADHD. Child Neuropsychology, 22, 618–626.

Hillstrom AP, & Yantis S (1994). Visual motion and attentional capture. Perception & Psychophysics, 55, 399–411.

Hitch GJ, Halliday S, Schaafstal AM, & Schraagen JMC (1988). Visual working memory in young children. Memory and Cognition, 16, 120–132.

Hoffman LM, & Gillam R (2004). Verbal and spatial information processing constraints in children with specific language impairment. Journal of Speech, Language, & Hearing Research, 47, 114.

Hudec KL, Alderson RM, Patros C, Lea S, Tarle S, & Kasper L (2015). Hyperactivity in boys with ADHD: The role of executive and non-executive functions. Res Dev Disabil, 45–46, 103–109.

JASP Team. (2018). JASP (Version 0.9.2 software).

Kass RE & Raftery AE (1995). Bayes factors. J American Statistical Association, 90, 773–795.

Kline RB (2016). Principles and practice of structural equation modeling. (4th Ed.). NY: Guilford.

Kofler MJ, Harmon S, Aduen P, Day T, …Sarver D (2018). Neurocognitive and behavioral predictors of social problems in ADHD: A bayesian framework. Neuropsychology, 32, 344–355.

Kofler MJ, Raiker J, Sarver D, Wells E, & Soto E (2016). Is hyperactivity ubiquitous in ADHD or dependent on environmental demands?Evidence from meta-analysis.Clin Psychol Rev,46,12–24.

Kofler MJ, Sarver D, Harmon S, Moltisanti A, Aduen P, Soto E, & Ferretti N (2018). Working memory and organizational skills problems in ADHD. J Child Psychology & Psychiatry, 59, 57.

Lahey BB, Pelham WE, Stein MA, Loney JAN, Trapani C, ... & Gold E (1998). Validity of DSM-IV ADHD for younger children. J Am Acad Child Adoles Psychiatry, 37, 695–702.

Marsh JE, Hughes RW, & Jones DM (2009). Interference by process, not content, determines semantic auditory distraction. Cognition, 110, 23–38.

Martel MM, Roberts B, Gremillion M, Von Eye A, & Nigg JT (2011). External validation of bifactor model of ADHD: Explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. J Abnormal Child Psychology, 39, 1111.

Martinez-Badía J, & Martinez-Raga J (2015). Who says this is a modern disorder? The early history of attention deficit hyperactivity disorder. World Journal of Psychiatry, 5, 379.

Medina JA, Netto TLB, Muszkat M, Medina AC, Botter D, Orbetelli R,,, Miranda MC (2010). Exercise impact on sustained attention of ADHD children, methylphenidate effects. ADHD Attention Deficit Hyperactivity Disorder, 2, 49–58.

Muthen LK, & Muthen BO (2012). Mplus user’s guide (7th ed.). Los Angeles: Muthen & Muthen.

Napolitano AC, & Sloutsky VM (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. Child Development, 75, 1850–1870.

Nee DE, Brown JW, Askren MK, German MG, Demiralp E, Krawitz A, & Jonides J (2013). A meta-analysis of executive components of working memory. Cerebral Cortex, 23, 264–282.

Nichols J, Shoulberg E, Garner A, Hoza B..Arnold L (2017). Exploration of the factor structure of ADHD in adolescence through self, parent, and teacher reports. J Abnorm Child Psychol, 45, 625.

Nigg JT, Blaskey LG, Huang-Pollock CL, & Rappley MD (2002). Neuropsychological executive functions and DSM-IV ADHD subtypes. J Am Acad Child Adoles Psychiatry, 41, 59–66.

Nunnally JC, & Bernstein IH (1994). Psychological theory. NY: MacGraw-Hill.

Patros CHG, Alderson RM, Hudec KL, Tarle SJ, & Lea SE (2017). Hyperactivity in boys with ADHD: The influence of underlying visuospatial working memory and self-control processes. Journal of Experimental Child Psychology, 154, 1–12.

Pelham WE, Fabiano GA, & Massetti GM (2005). Evidence-based assessment of ADHD in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34, 449–476.

Platzman KA, Stoy MR, Brown RT, Coles CD, Smith IE, & Falek A (1992). Review of observational methods in ADHD: Implications for diagnosis. School Psychology Quarterly, 7, 155.

Polanczyk G, De Lima M, Horta B, Biederman J, & Rohde L (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. American J Psychiatry,164,942.

Polanczyk G, Willcutt E, Salum G,…& Rohde L (2014). ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. Int J Epidemiol, 43, 434–442.

Pontifex MB, Parks AC, Henning DA, & Kamijo K (2015). Single bouts of exercise selectively sustain attentional processes. Psychophysiology, 52(5), 618–625.

Pontifex MB, Saliba BJ, Raine LB, Picchietti DL, & Hillman CH (2013). Exercise improves behavioral, neurocognitive, and scholastic performance in children with attention-deficit/ hyperactivity disorder. Journal of Pediatrics,

Postle BR, D'Esposito M, & Corkin S (2005). Effects of verbal and nonverbal interference on spatial and objective visual working memory. Memory & Cognition, 33, 203–212.

Rapport M, Bolden J, Kofler M, Sarver D, & Raiker J (2009). Hyperactivity in ADHD: Ubiqui-tous core symptom or manifestation of working memory deficits? J Abnorm Ch Psychol,37, 521–34.

Rees G, Frith C, & Lavie N (2001). Processing of irrelevant visual motion during performance of an auditory attention task. Neuropsychologia, 39(9), 937–949.

Rodriguez A, Reise SP, & Haviland MG (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. Journal of Personality Assessment, 98, 223–237.

Rodriguez A, Reise SP, & Haviland MG (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods, 21, 137–150.

Sarver DE, Rapport MD, Kofler MJ, Raiker JS, & Friedman LM (2015). Hyperactivity in ADHD: Impairing deficit or compensatory behavior?, J Abnormal Child Psychology, 43, 1219–32.

Satorra A, & Bentler PM (2010). Ensuring positiveness of the scaled difference chi-square test-statistic. Psychometrika, 75(2), 243–248.

Shrout PE, & Fleiss JL (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420.

Smith AL, Hoza B, Linnea K, McQuade JD, Tomb M, Vaughn AJ, Shoulberg E, & Hook H (2013). Pilot physical activity intervention reduces severity of ADHD symptoms in young children. Journal of Attention Disorders, 17(1), 70–82.

Siegel S, & Castellan NJ (1988). The case of k related samples. Nonparametric statistics for behavioral sciences. New York: McGraw-Hill, 170–4.

Theeuwes J, Kramer AF, Hahn S, Irwin DE, & Zelinsky GJ (1999). Influence of attentional capture on oculomotor control. Journal of Experimental Psychology Human Perception and Performance, 25, 1595–1608.

Tryon WW, Pinto LP, & Morrison DF (1991). Reliability assessment of pedometer activity measurements. Journal of Psychopathology and Behavioral Assessment, 13, 27–44.

Ullebø AK, Breivik K, Gillberg C, Lundervold AJ, & Posserud MB (2012). The factor structure of ADHD in a general population of primary school children. Journal of Child Psychology and Psychiatry, 53, 927–936.

Verret C, Guay M, Berthiaumer C, Gardiner P, & Beliveau L (2012) A physical activity program improves behavior and cognitive function in children with ADHD: An exploratory study. Journal of Attention Disorders, 16(1), 71–80.

Watkins MW (2017). Omega v.2 [Computer software]. Phoenix AZ: Ed & Psych Associates.

Wetzel N, & Schröger E (2014). On the development of auditory distraction. PsyCh Journal,3,72–91.

Willcutt EG, Nigg JT, Pennington BF, Solanto MV,…Lahey BB (2012). Validity of DSM-IV ADHD symptom dimensions and subtypes. Journal of Abnormal Psychology, 121, 991–1010.

Zhao X, Page TF, Altszuler AR, Pelham WE, Kipp H, Gnagy EM,...& Macphee FL (2019). Family Burden of Raising a Child with ADHD. Journal of Abnormal Child Psychology, 1–12.

Author Manuscript

Author Manuscript

Author Manuscript

Public Health Significance

Excess physical movement (i.e., hyperactivity) appears to be a compensatory behavior that facilitates cognition and functional outcomes for children with ADHD. At the same time, hyperactivity/impulsivity is associated with a host of negative outcomes. This study suggests that these conflicting findings may be related, at least in part, to subjective assessment methods that lump excess physical movement with other, cognitive-perceptually distinct behaviors assessed under the ADHD 'hyperactivity/impulsivity' umbrella.

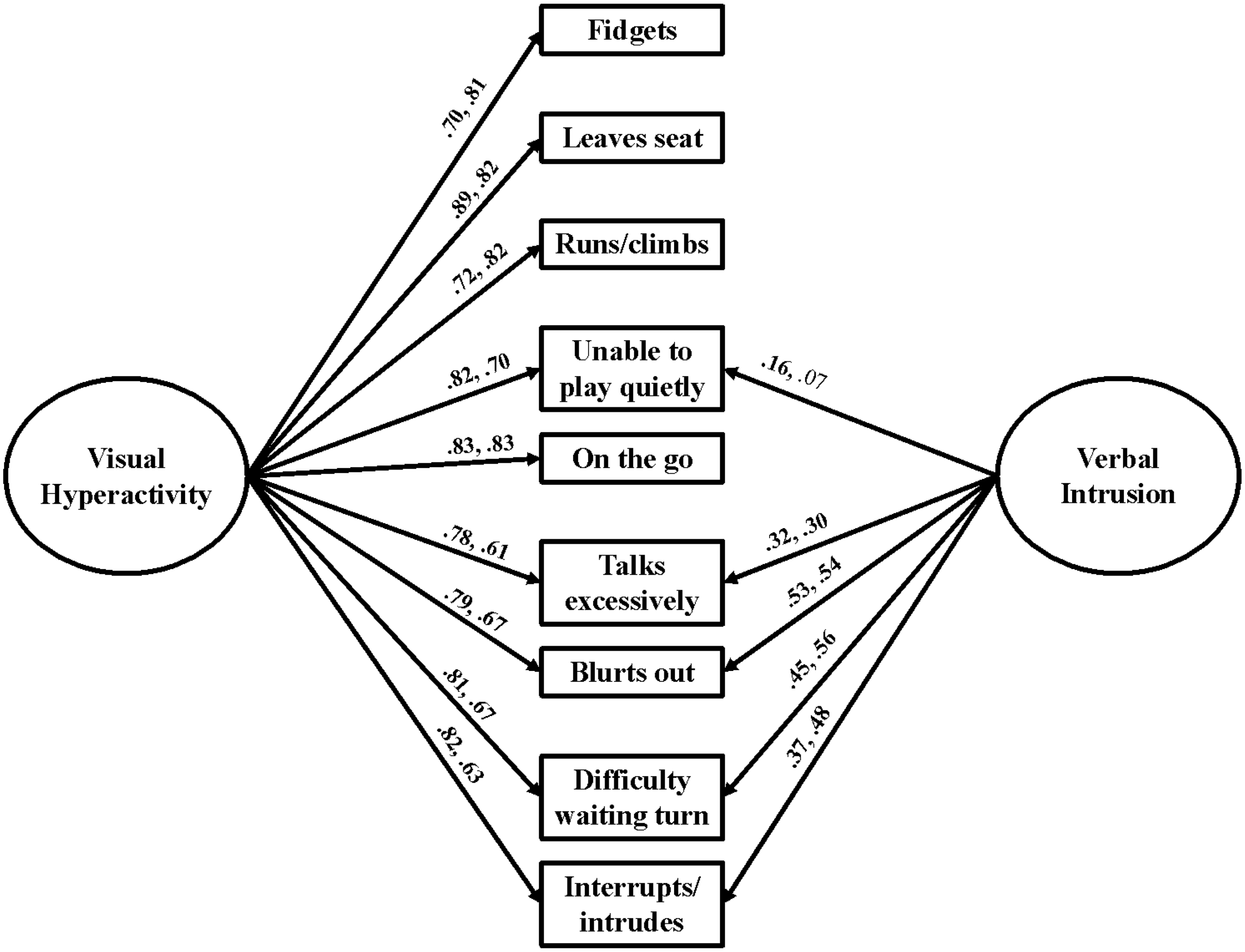

Figure 1.

Best-fitting visual hyperactivity and verbal intrusion model. Standardized factor loadings are shown for the teacher and parent models, respectively. Significant loadings (p<.05) are bolded.

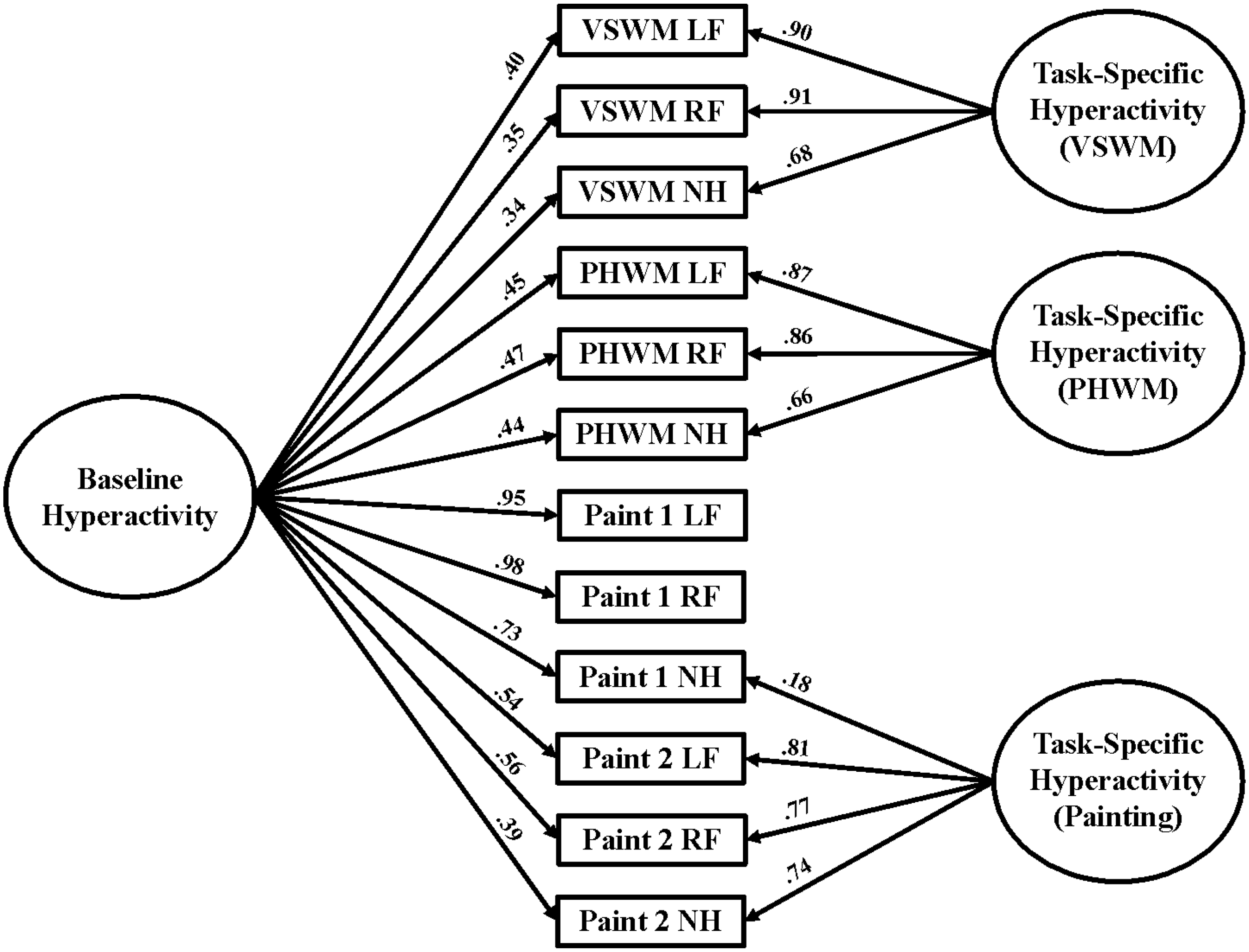

Figure 2.

Best-fitting actigraph hyperactivity model. Standardized factor loadings are shown. All loadings were significant at p<.003. LF = left foot actigraph, NH = non-dominant hand actigraph, Paint = beginning (1) and end (2) of session computerized painting activity, phonological working memory task, PHWM = phonological working memory task, RF = right foot actigraph, VSWM = visuospatial working memory task. Beginning of session left and right foot actigraphs served as the reference facets to define the general factor (baseline hyperactivity).

Table 1.

Sample and Demographic Variables

Variable	ADHD (N=82)		Non-ADHD (N=50)		Cohen's *d*	*p*	BF_{10}	BF_{01}
	M	***SD***	***M***	***SD***				
Gender (Boys/Girls)	55/27		30/20		--	.41		3.36
Ethnicity (AA/A/C/H/M)	11/0/60/7/4		5/3/28/10/3		--	.04		9.71
Age	10.06	1.44	10.79	1.54	−0.46	.01	5.59	
SES	47.49	11.57	49.59	12.24	−0.16	.32		3.35
FSIQ (Standard Scores)	103.01	15.43	108.02	10.35	−0.33	.04	1.22	
ADHD-5 Inattention (Raw Scores)								
Parent	19.32	6.09	14.26	8.14	0.69	< .001	257.62	
Teacher	16.73	6.61	10.38	7.73	0.86	< .001	8808.25	
ADHD-5 Hyperactivity/Impulsivity (Raw Scores)								
Parent	14.66	7.39	9.42	6.96	0.68	< .001	232.48	
Teacher	11.12	8.24	6.38	7.32	0.55	.001	26.65	
Actigraph Data (PIM)								
Paint 1 LF	10.63	10.92	7.33	8.71	0.33	.07		1.20
Paint 1 NH	14.96	11.14	11.64	11.39	0.30	.10		1.54
Paint 1 RF	11.95	12.29	7.76	11.08	0.35	.05	1.10	
Paint 2 LF	17.14	14.93	11.11	13.74	0.42	.02	2.11	
Paint 2 NH	23.31	19.25	15.80	15.39	0.42	.02	2.19	
Paint 2 RF	19.24	16.95	13.48	19.13	0.32	.08		1.22
PHWM LF	77.57	63.95	47.60	40.62	0.53	.004	9.46	
PHWM NH	92.72	53.34	70.92	48.56	0.42	.02	2.27	
PHWM RF	80.74	70.64	52.27	54.34	0.44	.02	2.73	
VSWM LF	54.25	49.55	26.48	24.14	0.67	< .001	77.11	
VSWM NH	79.63	50.01	44.17	27.80	0.83	< .001	1665.07	
VSWM RF	58.49	52.20	29.07	26.75	0.66	< .001	76.09	

Note. BF_{10} = Bayes Factor for the alternative hypothesis over the null hypothesis (values ≥ 3.0 indicate significant between-group differences; BF_{01} = 1/ BF_{10}). BASC = Behavior Assessment System for Children. Ethnicity: AA = African American, A = Asian, C = Caucasian Non-Hispanic, H = Hispanic, M = Multiracial. PIM = Proportional Integrating Measure (movement intensity). FSIQ = Full Scale Intelligence (WISC-V Short Form), LF = left foot actigraph, NH = non-dominant hand actigraph, Paint = beginning (1) and end (2) of session computerized painting activity, phonological working memory task, PHWM = phonological working memory task, RF = right foot actigraph, SES = Hollingshead socioeconomic status, VSWM = visuospatial working memory task.

Table 2.

Item-level judgments for hypothesized visual hyperactivity/verbal intrusion factor structure

Item	Physical Symptoms (Hyperactivity)	Auditory Symptoms (Verbal Intrusion)	Inter-judge Agreement (K=7)	Teacher Report M (SD)	Teacher Skewness (SE)	Teacher Kurtosis (SE)	Parent Report M (SD)	Parent Skewness (SE)	Parent Kurtosis (SE)
Fidgets/squirms	X		100%	1.68 (1.16)	−0.15 (0.21)	−1.46 (0.42)	1.86 (110)	−0.51 (0.21)	−1.08 (0.42)
Difficulty remaining seated	X		100%	1.13 (114)	0.56 (0.21)	−1.13 (0.42)	1.46 (116)	0.11 (0.21)	−1.43 (0.42)
Runs/climbs	X		100%	0.55 (0.89)	1.47 (0.21)	1.01 (0.42)	1.02 (107)	0.71 (0.21)	−0.78 (0.42)
Difficulty playing quietly		X	100%	0.88 (0.99)	0.82 (0.21)	−0.46 (0.42)	1.00 (0.97)	0.83 (0.21)	−0.18 (0.42)
On the go/driven by a motor	X		100%	0.96 (111)	0.69 (0.21)	−0.98 (0.42)	1.29 (117)	0.29 (0.21)	−1.40 (0.42)
Talks excessively		X	100%	1.24 (113)	0.42 (0.21)	−1.21 (0.42)	1.63 (108)	−0.06 (0.21)	−1.29 (0.42)
Blurts out answers		X	100%	1.02 (116)	0.68 (0.21)	−1.05 (0.42)	1.53 (103)	0.02 (0.21)	−1.14 (0.42)
Difficulty waiting turn[1]		X	71%	0.89 (104)	0.90 (0.21)	−0.44 (0.42)	1.32 (102)	0.24 (0.21)	−1.05 (0.42)
Interrupts people		X	89%	1.03 (111)	0.65 (0.21)	−0.98 (0.42)	1.64 (100)	0.02 (0.21)	−1.03 (0.42)
Internal consistency (current sample)									
Parent report	ω=.89, α=.89	ω=.88, α=.87							
Teacher report	ω=.87, α=.84	ω=.95, α=.95							

Note: In the bifactor s−1 model, all items load onto the general factor (hyperactivity), and only the verbal intrusion items load onto the specific factor (verbal intrusion). Obtained range for all teacher and parent items was 0–3 (maximum possible range = 0–3). Overall interjudge agreement (ICC) = .95, Fleiss's kappa = .85. Qualitatively, all items were categorized with >80% agreement as belonging to either the gross motor movement (visual hyperactivity) or auditory interruptions/intrusions (verbal intrusion) categories with one exception: item 8 'difficulty waiting turn' was rated as belonging to both categories by 71% of judges (Table 2); this item was also included in both categories in Gibbins et al. (2012). We resolved this discrepancy empirically by comparing models with vs. without the 'difficulty waiting turn' item loading onto the specific verbal intrusion factor. We retained the models that included this loading based on improved model fit for both the parent and teacher models (both $\Delta\chi^2$ (1) > 43.21, $p < .001$; differences in additional fit statistics were: ΔCFI=.04–.06, ΔTLI=.05–.10, ΔRMSEA= −.05 for both models, ΔAIC = −42 to −48, ΔBIC = −39 to −45).

[1] Judged to belong equally to both physical and auditory categories by 5 of 7 judges.

Table 3.

Measurement models: Teacher-reported hyperactivity

Model	CFI	TLI	RMSEA (90%CI)	AIC	BIC	General factor loadings	Specific factor loadings
DSM-5 hyperactivity/impulsivity single-factor	.93	.91	.15 (.12–.18)	2563	2615	.61–.92, all $p < .001$	--
DSM-IV hyperactivity/impulsivity correlated factors	.96	.94	.12 (.08–.15)	2535	2590	.65–.84, all $p < .001$	.91–.94, all $p < .001$
DSM-IV hyperactivity/impulsivity (bifactor s−1)	.96	.94	.12 (.08–.15)	2534	2595	.65–.87, all $p < .001$	.26–.45, all $p < .004$
Visual hyperactivity/verbal intrusion correlated factors	.96	.94	.12 (.08–.15)	2535	2590	.70–.83, all p<.001	.81–.91, all p<.001
Visual hyperactivity/verbal intrusion (bifactor s−1)	**.97**	**.95**	**.10 (.07–.14)**	**2523**	**2589**	**.70–.89, all $p < .001$**	**.32–.45, all $p < .001$ ***

Note: Preferred model in **bold**. For bifactor models, hyperactivity is the general factor (items 1–9 as indictors) and impulsivity (items 7–9) or verbal intrusion (items 7, 4, 6, 7, 8, 9) are the specific factors; item 7 (blurts out) served as the reference variable for both bifactor models for maximal comparability across models. For the correlated factors models, loadings for visual hyperactivity are reported under the "general" column and loadings for verbal intrusion are reported under the "specific" column for readability (the correlated factor model does not impose a hierarchical structure).

* Item 4 (difficulty playing quietly) loaded .16 ($p = .027$)

Table 4.

Measurement models: Parent-reported hyperactivity

Model	CFI	TLI	RMSEA (90%CI)	AIC	BIC	General factor loadings	Specific factor loadings
DSM-5 hyperactivity/impulsivity single-factor	.87	.82	.17 (.14–.20)	2866	2918	.67–.82, all $p < .001$	--
DSM-IV hyperactivity/impulsivity correlated factors	.94	.91	.12 (.09–.15)	2813	2886	.65–.82, all $p < .001$	.80–.87, all $p < .001$
DSM-IV hyperactivity/impulsivity (bifactor s−1)	.94	.90	.13 (.09–.16)	2816	2876	.65–.82, all $p < .001$	.45–.57, all $p < .001$
Visual hyperactivity/verbal intrusion correlated factors	.93	.90	.13 (.10–.16)	2821	2876	.81–.85, all p<.001	.65–.85, all p<.001
Visual hyperactivity/verbal intrusion (bifactor s−1)	**.95**	**.92**	**.12 (.08–.15)**	**2808**	**2874**	**.61–.83, all $p < .001$**	**.30–.56, all $p < .001$** *

Note: Preferred model in **bold**. For bifactor models, hyperactivity is the general factor (items 1–9 as indictors) and impulsivity (items 7–9) or verbal intrusion (items 7, 4, 6, 7, 8, 9) are the specific factors; item 7 (blurts out) served as the reference variable for both bifactor models for maximal comparability across models. For the correlated factors models, loadings for visual hyperactivity are reported under the “general” column and loadings for verbal intrusion are reported under the “specific” column for readability (the correlated factor model does not impose a hierarchical structure).

* Item 4 (difficulty playing quietly) failed to load: .07 ($p = .44$)

Table 5.

Measurement models: Actigraph-measured hyperactivity

Model	CFI	TLI	RMSEA (90%CI)	AIC	BIC	General factor loadings	Specific factor loadings
Hyperactivity single-factor	.42	.29	.39 (.37–.41)	14029	14098	.46–.95, all $p < .001$	--
Hyperactivity general/task-specific (bifactor s−1)	**.94**	**.91**	**.14 (.12–.17)**	**13071**	**13171**	**.34–.99, all $p < .001$**	**.63–.91, all $p < .001$** *

Note: Preferred model in **bold**. For bifactor models, hyperactivity is the general factor (all actigraphs indicators for all tasks as indictors) and there is a specific factor for each task (baseline, phonological working memory, visuospatial working memory); the left and right foot actigraphs for the beginning of session baseline activity served as the index variables (i.e., loaded on the general factor but not on a specific factor).

* the non-dominant hand actigraph for the beginning of session baseline loaded .18, $p = .003$.

Table 6.

Structural model: Associations between teacher ratings and actigraph-measured hyperactivity (correlated bifactor s−1 models)

Model	CFI	TLI	RMSEA (90%CI)	AIC	BIC	General factor loadings	Specific factor loadings
Visual hyperactivity/verbal intrusion bifactor s−1 correlated with actigraph bifactor s−1	.95	.93	.08 (.07–.10)	15579	15766	Hyp: .70–.88, all $p < .001$	Verbal: .31–.52, all $p < .001$ *
						Act: .39–.98, all $p < .001$	PHWM: .66–.87, all $p < .001$
							VSWM: .68–.91, all $p < .001$
							Paint: .74–.82, all $p < .001$ **

Note: For bifactor models of informant ratings, hyperactivity is the general factor (items 1–9 as indictors) and impulsivity (items 7–9) or verbal intrusion (items 7, 4, 6, 7, 8, 9) are the specific factors; item 7 (blurts out) served as the reference variable for both bifactor models for maximal comparability across models. For bifactor model of actigraph-measured hyperactivity, baseline gross motor activity is the general factor (all actigraph datapoints as indicators) and task-specific hyperactivity during the phonological working memory (PHWM), visuospatial working memory (VSWM), and Paint tasks are the specific factors; the left and right foot actigraphs for the beginning of session baseline activity served as the index variables (i.e., loaded on the general factor but not on a specific factor). Act = actigraph general factor (baseline hyperactivity); Hyp = hyperactivity ratings general factor; Imp = impulsivity specific factor; Paint = task-specific hyperactivity during the beginning and end of session computer paint activity; PHWM = task-specific hyperactivity during the phonological working memory task; Verbal = verbal intrusion specific factor; VSWM = task-specific hyperactivity during the visuospatial working memory task.

* Item 4 (difficulty playing quietly) loaded .15 ($p = .038$)

** The non-dominant hand actigraph for the beginning of session baseline loaded .18, $p = .003$.

Table 7.

Structural model: Associations between parent ratings and actigraph-measured hyperactivity (correlated bifactor s−1 models)

Model	CFI	TLI	RMSEA (90%CI)	AIC	BIC	General factor loadings	Specific factor loadings
Visual hyperactivity/verbal intrusion bifactor s−1 correlated with actigraph bifactor s−1	.94	.92	.09 (.07–.10)	15859	16046	Hyp: .61–,83, all $p < .001$	Verbal: .31–.52, all $p < .003$ *
						Act: .34–.98, all $p < .001$	PHWM: .66–.87, all $p < .001$
							VSWM: .68–.91, all $p < .001$
							Paint: .74–.82, all $p < .001$ **

Note: For bifactor models of informant ratings, hyperactivity is the general factor (items 1–9 as indictors) and impulsivity (items 7–9) or verbal intrusion (items 7, 4, 6, 7, 8, 9) are the specific factors; item 7 (blurts out) served as the reference variable for both bifactor models for maximal comparability across models. For bifactor model of actigraph-measured hyperactivity, baseline gross motor activity is the general factor (all actigraph datapoints as indicators) and task-specific hyperactivity during the phonological working memory (PHWM), visuospatial working memory (VSWM), and Paint tasks are the specific factors; the left and right foot actigraphs for the beginning of session baseline activity served as the index variables (i.e., loaded on the general factor but not on a specific factor). Act = actigraph general factor (baseline hyperactivity); Hyp = hyperactivity ratings general factor; Imp = impulsivity specific factor; Paint = task-specific hyperactivity during the beginning and end of session computer paint activity; PHWM = task-specific hyperactivity during the phonological working memory task; Verbal = verbal intrusion specific factor; VSWM = task-specific hyperactivity during the visuospatial working memory task.

* Item 4 (difficulty playing quietly) failed to load: .08 ($p = .38$)

** The non-dominant hand actigraph for the beginning of session baseline loaded .18, $p = .003$.