HHS Public Access

Author manuscript

J Psychopathol Behav Assess. Author manuscript.

Ecological Validity of Clinic-Based Actigraphy for Assessing Hyperactivity in Clinically Evaluated Children with and without ADHD

Michael J. Kofler, Ph.D.[1], **Elia F. Soto, Ph.D.**[2], **Mark D. Rapport, Ph.D.**[3], **Samuel Whitehead**[3]

[1]Florida State University, Department of Psychology

[2]Louisiana State University, Department of Psychology

[3]University of Central Florida, Department of Psychology

Abstract

Though often conflated with face validity, ecological validity refers to the degree that a test or measure predicts real-world behavior/functioning. The current study leveraged two independent samples to provide a critical evaluation of the extent to which clinic-based actigraphy demonstrates ecological validity evidence relative to parent- and teacher-reported hyperactivity ratings. Further, across both samples we evaluated the extent to which the ecological validity evidence for these mechanical measures of hyperactivity varies as a function of the task children are completing while their movement is assessed objectively (low vs. high cognitive demands). Across two independent samples comprising clinically-evaluated children with and without ADHD (*Ns*=88, 184; M_{ages}=9.2, 10.4; 6%, 33% girls; 68%, 70% White Non-Hispanic), latent path models indicated that clinic-based actigraphy during visuospatial working memory testing (high cognitive demands) demonstrated significant associations with both parent- and teacher-rated hyperactivity that were indistinguishable (p>.05) from parent and teacher ratings's associations with each other in both sample 1 (r=.57) and sample 2 (r=.35; all p<.001). Actigraphy during baseline (low cognitive demand) conditions also uniquely predicted hyperactivity at home and school in both samples (all p<.001), albeit with a less consistent yet robust pattern relative to parent/teacher associations. In both samples, actigraphy showed strong test-retest reliability over 2–4 weeks across clinic-based tasks with high cognitive demands (r=.61-.93) and high concurrent validity across tasks with high vs. low cognitive demands (r=.35-.61; all p<.007). This pattern supports the ecological validity of clinic-based actigraphy during working memory testing, which predicts real-world behavior at home and school just as well as parent perceptions of hyperactivity at home predict teacher perceptions of hyperactivity at school (and vice versa).

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

Author Contributions: The authors contributed substantively, collaboratively, and interactively in the production of this manuscript. The authors decline to parcel credit for specific tasks to specific people because we feel this promotes competition rather than collaboration. All authors read and approved the final manuscript.

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

actigraph; hyperactivity; ecological validity; ADHD

The neurodevelopmental disorder ADHD – attention-deficit/hyperactivity disorder – affects an estimated 5–7% of school-age children worldwide (Polanczyk et al., 2014) and as high as 10.5% are diagnosed with ADHD in the United States based on a recent National Survey of Children's Health (Danielson et al., 2024). Clinical subtypes of the disorder are designated based on symptom presentations within two primary domains (inattention, hyperactivity-impulsivity; American Psychiatric Association, 2022), with the hyperactivity subdomain being of particular relevance due to its central role in historical nomenclatures of the disorder, and as a harbinger of both positive and negative associations and outcomes (for reviews, see Kofler et al., 2016; 2020; Rapport et al., 2009). It is also chief among teacher and parent complaints, and a motivating force for initiating pharmacological, psychosocial, and/or educational interventions (Barkley, 2016; DuPaul & Stoner, 2014; Fabiano et al., 2021).

Assessing hyperactive behavior in children with ADHD has involved a wide array of methodological approaches, ranging from analogue measures (Barkley, 1991) and direct observations (Rapport, Kofler et al., 2009; Sarver et al., 2015) to more objective, mechanical measures such as pedometers (Plomin & Foch, 1981), infra-red motion imaging (Lee et al., 2021; Murillo et al., 2015), piezoelectric chair movements (Chang et al., 2023), and actigraphs (Alderson et al, 2012; De Crescenzo et al., 2016; Dekkers et al., 2021; Kofler et al., 2020; Murillo et al., 2015; Porrino & Rapoport, 1983; Rapport, Bolden et al., 2009). Despite the sophistication and improved reliability of these latter measures, parent/teacher ratings remain the gold standard of quantifying excessive motor movement in children with ADHD for clinical/diagnostic purposes and monitoring treatment related outcomes (Barkley, 2019; DuPaul & Stoner, 2014; Kofler et al., 2020).

The preference for and popularity of using parent/teacher ratings to assess hyperactivity reflects the economic utility of obtaining, administering, and scoring the instruments, and their oft argued superior ecological validity relative to objective mechanical measures such as actigraphs (Barkley, 1991; Dane, Schachar, & Tannock, 2000; Halperin et al., 1993). Hyperactivity rating scales are obtained at minimal or no cost, can be completed quickly and by multiple informants to assess the breadth, severity, and situational nature of excess motor activity relative to established norms, and interpreted readily by mental health, educational, and medical professionals. Proponents embracing arguments concerning the enhanced ecological validity of rating scales over clinic-based options often cite evidence regarding the low to moderate correlations between informant ratings and objective actigraphy (r=.32-.58; Rapport, Kofler, & Himmerich, 2006). Further, prior evidence suggests that unstructured, qualitative assessment of a child's activity level in a clinic setting such as a physician's office corresponds weakly (correctly identifies approximately 20% of children meeting full diagnostic criteria) with their behavior in more naturalistic settings such as the home or at school (Sleator & Ullmann, 1981).

Ecological validity refers to the extent to which a test or measure predicts real-world functioning (Andrade, 2018). Though often conflated with face validity (Suchy et al., 2024a, 2024b), ecological validity is assessed quantitatively as the correlation between a test/measure and an appropriate measure in naturalistic practice or in everyday life (Andrade, 2018). In that context, conclusions that clinic-based actigraphy possesses lower than desired ecological validity likely reflects at least three factors – (a) the conflation of ecological validity with face validity (Suchy et al., 2024a, 2024b), resulting in qualitative rather than quantitative assessments of ecological validity despite evidence that face validity likely does not enhance ecological validity (Ziemnik & Suchy, 2019); (b) the conflation of hyperactivity and impulsivity symptoms contained on most rating scales when assessing the ecological validity of high-precision measures of hyperactivity such as actigraphs; and (c) the common practice of measuring children's motor movement via actigraphy primarily during repetitive/monotonous activities or tasks with low cognitive demands that evoke comparatively lower levels of movement than those observed in educational settings (for review, see Kofler et al., 2016).

The initial issue – conflation of ecological validity (ability to *predict* real-world outcomes) and face validity (*resemblance* to real-world settings/situations) – may explain why critiques of actigraphy's ecological validity have generally not compared associations between clinic-based actigraphy and 'real world' settings such as home vs. school (Suchy et al., 2024a, 2024b). Indeed, there seems to be an implicit assumption that clinic-based actigraphy is inherently not ecologically valid, as evidenced by studies explicitly examining actigraphy's ecological validity overwhelmingly using these devices in real-world settings (e.g., home; Kofler et al., 2016). To that end, meta-analytic evidence indicates that cross-informant ratings of excess motor activity (near ubiquitously measured via scales that combine hyperactivity and impulsivity) correlate only moderately with each other at r=.10-.50 (e.g., Saffer et al., 2021; Takeda et al., 2020). Given that highly similar magnitude associations are reported between clinic-based actigraphy and parent/teacher ratings (r=.32-.58; Rapport et al., 2006), one might already argue for the ecological validity of clinic-based actigraphy. That is, clinic-based actigraphy appears to predict real-world behavior just as well as parent reports of hyperactivity at home predict teacher reports of hyperactivity at school (and vice versa). However, to our knowledge, no study to date has used latent modeling to compare these estimates head-to-head within the same sample – let alone replicated these estimates across independent samples.

The second factor – imprecise measurement of excess motor movement (hyperactivity) – occurs because rating scale items nearly always reflect a broader range of behavior beyond gross motor movement. For example, the most commonly used hyperactivity scales include all nine DSM-5 (APA, 2022) hyperactivity-impulsivity symptoms. The conflation of hyperactivity and impulsivity items is well documented (DuPaul et al., 2015), and elucidated recently using structural modeling based on parent and teacher ratings of DSM-5 hyperactivity/impulsivity symptoms (Kofler et al., 2020). In that study, a bifactor model emerged for both informants consisting of two uncorrelated factors – *visual hyperactivity* and *verbal intrusiveness* – and showed superior fit relative to a correlated factors model, as well as strong convergent validity between the visual hyperactivity factor and collected actigraph data (Kofler et al., 2020). 'Visual hyperactivity' refers to excess

physical movement that is observable to others, and is reflected by four of the nine DSM hyperactivity-impulsivity items (fidgets, leaves seat, runs/climbs, on the go), The term 'verbal intrusiveness' was preferred over the traditional 'impulsivity' label for the remaining items because these items describe excess vocalizations or other noises that gain automatic access to other children's phonological loops, where they necessarily compete with verbal/auditory information currently held in mind (e.g., can't play quietly, talks excessively, blurts out, interrupts; Kofler et al., 2020).[1]

The third factor noted above – the practice of measuring children's motor movement via actigraphy under monotonous or low cognitive demand conditions (see Kofler et al., 2016 for review) – is also a plausible candidate that contributes to the ecological validity conundrum. Children with ADHD, as well as same age typically developing (TD) children, move significantly more when engaged in high relative to low cognitive demand tasks (for review see Kofler et al., 2016). In contrast, children with and without ADHD are largely indistinguishable from one another under low cognitive demand task conditions (Hudec et al., 2015; Kofler et al., 2016; Patros et al., 2017; Rapport et al., 2008; Sarver et al., 2015; cf. Soto et al., 2024), including fast paced, stimulating activities (e.g., action movies) that require rapid visual processing but place minimal demands on higher level executive functions such as working memory (Orban, Rapport, Friedman, Eckrich, & Kofler, 2017). Of particular relevance is the replicated finding that children with ADHD move at nearly twice the rate of TD children under high cognitive demand conditions in analogue settings (Dekkers et al., 2021; Patros et al., 2017; Rapport et al., 2008). These findings render task demands as a critical parameter to consider when attempting to approximate levels of excess gross motor activity observed in the classroom that occur as children engage in instructional activities (e.g., reading, mathematics, science, writing) requiring moderate to high levels of cognitive processing throughout the day (Kofler et al., 2019; Spiegel, Goodrich, Morris, Osborne, & Lonigan, 2021).

Collectively, the conflation of ecological validity with face validity, inclusion of non-movement related items in parent/teacher hyperactivity rating scales, and quantification of movement via actigraphs under monotonous/low cognitive demand conditions represent high-profile candidates responsible for perceptions in the field that clinic-based measurement of children's activity level is not ecologically valid. Beyond the definitional issue, the Kofler et al. (2020) study addressed the second of these concerns, and provided robust empirical support for using DSM-5 *visual hyperactivity* symptoms alone, rather than a combination of hyperactivity and impulsivity (*verbal intrusiveness*) items, to estimate hyperactivity in children. A complementary study, wherein parent/teacher ratings of visually observed hyperactivity are contrasted with actigraph measured hyperactivity under low and high cognitive demand conditions warrants scrutiny to address the remaining ecological validity issue, and is the focus of the present investigation.

[1]The final item, difficulty waiting turn (e.g., while waiting in line), could conceptually reflect either overt movement/hyperactivity (e.g., getting out of line) or unwanted vocalizations (e.g., making noises, humming instead of waiting quietly in line). Ratings on this item appear to be driven by excess vocalizations while waiting in line based on improved model fit when this item was allowed to load on the verbal intrusion factor (Kofler et al., 2020).

To address the issues described above, latent path modeling was used to test the extent to which clinic-based measurement of hyperactivity (actigraphy) demonstrates similar levels of ecological validity as cross-setting (school vs. home) ratings. This involved creating and testing latent associations among 4 measures of hyperactivity: parent report, teacher report, actigraphy during tasks with low cognitive demands, and actigraphy during working memory tests with high cognitive demands. This model was then replicated in an independent sample. In other words, in two independent samples, we asked whether clinic-based measurement of hyperactivity was comparable to parent ratings of hyperactivity at home in terms of their ecological validity for predicting hyperactivity at school (and vice versa). We expected the results to improve our understanding of how best to optimize actigraph-measured hyperactivity in clinical settings to mirror children's hyperactivity as it is observed within home and classroom settings. Looking ahead, we predict that the convergence of the two methods (ratings, actigraphy) will provide clinicians with a more comprehensive armamentarium to complement their diagnostic decision making and may prove useful for treatment monitoring while reducing demands on teachers and parents and eliminating confounds associated with rater biases such as recency and halo effects.

Methods

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Data were analyzed using JASP 0.19.1 (JASP Team, 2024). The study was not pre-registered; however, all measure inclusion/exclusion decisions and analytic plans were made *a priori*, prior to accessing the data except as clearly noted. Anonymized data/code and results output are available from the authors upon reasonable request. Working memory performance data were reported for the current samples in Orban et al. (2018) (Sample 1) and Gaye et al. (2025) (Sample 2). Actigraph data for the replication sample (Sample 2) were reported in Soto et al. (2024) as part of a larger experiment probing specific cognitive and non-cognitive processes that do and do not provoke hyperactive behavior in children with vs. without ADHD; associations between actigraph scores and parent/teacher behavior ratings have not been reported previously for either sample. Descriptions in the Methods section are reproduced/adapted from our standard research/clinic recruitment and testing protocols licensed under CC BY 4.0.

Participants

The current study leveraged two independent samples collected from different universities in the Southeastern United States. Children were recruited/referred at each site to a university-based children's learning clinic through community resources for participation in larger studies of the neurocognitive mechanisms underlying pediatric attention/behavioral problems. Institutional Review Board approval was separately obtained/maintained at each site; all parents and children gave informed consent/assent. A subset of Sample 2's ADHD group participated in a clinical trial of a computerized psychosocial intervention after completing the testing reported herein; only pretreatment data is used in the current study.

Sample 1 comprised 88 children (5 girls, 83 boys), aged 8 to 12 years (M=9.63, SD=1.28). Sample 1's race/ethnicity was 59 White Non-Hispanic (67.8%), 4 Black or African American (4.6%), 16 Hispanic or Latino (18.4%), 6 multiracial (6.9%), and 2 Asian (2.3%) children. Sample 2 comprised 184 children (61 girls, 123 boys), aged 8 to 13 years (M=10.40, SD=1.50). Sample race/ethnicity was mixed with 129 White Non-Hispanic (70.1%), 24 Black or African American (13.0%), 13 Hispanic or Latino (7.1%), 17 multiracial children (9.3%), and 1 Asian (0.5%) child (Tables 1a and 1b).

Group Assignment

Children and caregivers at both sites completed a comprehensive psychoeducational evaluation that included detailed parent semi-structured clinical interviewing using the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS; Kaufman et al., 1997). The K-SADS (2013 Update for DSM-5) allows differential diagnosis according to symptom onset, course, duration, quantity, severity, and impairment in children and adolescents based on DSM-5 criteria. Its psychometric properties are well established, including inter-rater agreement of .93 to 1.00, test-retest reliability of .63 to 1.00, and concurrent (criterion) validity between the K-SADS and psychometrically established parent rating scales (Kaufman et al., 1997). K-SADS interviews were supplemented with parent and teacher rating scales (Sample 1: Child Behavior Checklist/ Teacher Report Form, Child Symptom Inventory-IV; Achenbach & Rescorla, 2001), Gadow et al., 2004; Sample 2: ADHD Rating Scale-4/5, Behavior Assessment System for Children-2/3; DuPaul et al., 2016; Reynolds & Kamphaus, 2004). A psychoeducational report was provided to parents; children selected a small toy (<$5) from a prize box after each session; teacher received a $20 gift card.

The ADHD group at each site comprised children who met all of the following criteria: (1) DSM-5 diagnosis of ADHD by the directing clinical psychologist and multidisciplinary treatment team based on KSADS and differential diagnosis considering all available clinical information indicating onset, course, duration, and severity of ADHD symptoms consistent with the ADHD neurodevelopmental syndrome; (2) borderline/clinical elevations on at least one parent and one teacher ADHD subscale (i.e., >90th percentile); and (3) current impairment based on parent report (K-SADS). Children with all ADHD current presentation specifiers were eligible given the instability of ADHD subtypes (Lahey et al., 2005; Valo & Tannock, 2010; Willcutt et al., 2012): Sample 1: 90.0% Combined, 8.0% Inattentive, 2.0% Hyperactive/Impulsive; Sample 2: 68.9% Combined, 29.4% Inattentive, 1.7% Hyperactive/ Impulsive Presentations.

Children were excluded from both samples for gross neurological, sensory, or motor impairment that would preclude valid test administration; nonstimulant medications that could not be withheld for testing; or history of seizure disorder, autism spectrum disorder, psychosis, or intellectual disability. Psychostimulants were withheld >24-hours for neurocognitive testing. Additional exclusionary criteria for Sample 1 included additional clinical diagnoses beyond oppositional defiant disorder (ODD; 6.0%) and specific learning disorders (SLD; 10.2%).

For Sample 2, additional clinical diagnoses were not exclusionary. The standard assessment battery included norm-referenced child internalizing disorder screeners, and additional standardized measures were administered clinically as needed to inform differential diagnosis and accurate assessment of comorbidities (e.g., child clinical interviews, additional testing). Co-occurring conditions in Sample 2 included ODD[2] (11.8%), depressive disorders (3.4%), and anxiety disorders (28.6%). Children with ADHD were screened for SLD in reading (12.6%) and math (6.7%) defined by score(s) >1.5 SD below age-norms on one or more subtest(s) of the Academic Skills Battery of the Kaufman Test of Educational Achievement, Third Edition (Kaufman, 2014).

Sample 1 also included a typically developing control group of 38 children (2 girls) who did not meet ADHD criteria, completed the same comprehensive evaluation as the ADHD group, had normal developmental histories, and scored in the non-clinical range across the standardized parent and teacher ratings described above. For Sample 2, to account for the inclusion of co-occurring conditions in the ADHD group, the Non-ADHD group comprised 65 consecutive case-control referrals (25 girls) who did not meet ADHD criteria and included both neurotypical children and children with psychiatric disorders other than ADHD. Neurotypical children (61.5%) had normal developmental histories and nonclinical parent/teacher ratings, were recruited through community resources, and completed the same evaluation as clinically referred cases. Clinically referred and evaluated children who did not meet ADHD criteria were also included in the Non-ADHD group, including diagnoses of anxiety (29.2%) and depressive (9.2%) disorders. The ADHD and Non-ADHD groups did not differ in the proportion of children diagnosed with anxiety or depression, though rates of ODD were higher in the ADHD group as expected (Soto et al., 2024).

Procedures

Working memory testing occurred as part of larger batteries that involved 4 (Sample 1) or 2 (Sample 2) sessions of approximately 3 hours each. Procedures were identical across sites. All tasks were counterbalanced across testing sessions to minimize order effects. Children received brief breaks after each task and preset longer breaks every 2–3 tasks to minimize fatigue. Children were seated in a caster-wheel swivel chair (Rapport et al., 2009). Performance was monitored at all times by the examiner, who was stationed just out of the child's view to provide a structured environment while minimizing improvements related to examiner demand characteristics (Gomez & Sanson, 1994).

Working Memory

The Rapport et al. (2009) computerized working memory tests and instructions are identical to those described in Kofler et al. (2018b). Reliability and validity evidence includes high internal consistency (α = .81-.97) and 1–3-week test-retest reliability (.76-.90; Kofler et al., 2019; Sarver et al., 2015), and expected relations with criterion working memory complex span (r = .69) and updating tasks (r = .61) (Wells et al., 2018). Five practice trials were administered before each task (80% correct required). Sample 1 completed

[2]As recommended in the K-SADS, oppositional defiant disorder was diagnosed clinically only with evidence of multi-informant/multi-setting symptoms. ODD comorbidity is 34% in the ADHD group and 17% in the Non-ADHD group based on parent-reported symptom counts.

24 trials at each set size (3–6), with one counterbalanced set size per task administered on each of the four testing days. Sample 2 completed 6 trials at each set size (3–6) that were administered within single tasks in randomized/unpredictable order (separate counterbalanced phonological and visuospatial tasks; 3–6 stimuli/trial; 24 total trials per task; Kofler et al., 2016).

Phonological working memory. Children were presented a series of jumbled numbers and a letter (1 stimuli/s). The letter was never presented first or last to minimize primacy/recency effects and was counterbalanced to appear equally in the other serial positions. Children reordered and recalled the numbers from least to greatest, and said the letter last (e.g., 4H62 is correctly recalled as 246H).

Visuospatial working memory. Children were shown nine squares arranged in three offset vertical columns. A series of 2.5 cm dots were presented sequentially (1 stimuli/s); no two dots appeared in the same square on a given trial. All dots were black except one red dot that never appeared first or last to minimize primacy/recency effects. Children reordered the dot locations (black dots in serial order, red dot last) and responded on a modified keyboard.

Baseline Control Conditions

Children used Microsoft® Paint for five consecutive minutes at the beginning and end of each research session. The Paint program served as pre- and post-conditions to assess and control for potential within-day fluctuations in activity level (e.g., fatigue effects), and as an in-seat condition with marginal cognitive demands for comparative purposes. Children sat in the same chair and interacted with the same computer used for the cognitive task battery (i.e., the Paint program allows children to draw/paint anything they like on the monitor using a variety of interactive tools).

Objectively-measured Hyperactivity

Actigraph. 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. Higher scores indicate greater intensity of movement (proportional integrating measure/PIM mode). Sample 1 wore Motionlogger MicroMini actigraphs, whereas Sample 2 wore the Micro Motionlogger model of actigraphs (AMI/Ambulatory Monitoring, Inc., 2014a, 2014b). These actigraphs provide data on total movement intensity per task and mean movement intensity per second during each task, respectively.

Dependent variables. Following Rapport et al. (2009), we computed Total Hyperactivity Scores (THS) by summing activity level across the three actigraph sites (2 ankles and 1 non-dominant hand) to index total movement, separately for each task. This approach has the advantage of increasing power while providing a broad sampling of children’s activity

level (Rapport et al. 2009; Soto et al., 2024). Separate THS scores were computed for each of the working memory tasks, as well as separately for the Paint beginning and end tasks. To maximize comparability across samples, Paint THS scores from the first two testing sessions were used.

Global Intellectual Functioning (IQ) and Socioeconomic Status (SES)

All children were administered the Wechsler Intelligence Scale for Children (WISC-V) (sample 1) or the WISC-V Verbal Comprehension Index (VCI; sample 2; Wechsler, 2014). Hollingshead SES was estimated based on caregiver(s)' education and occupation (Cirino et al., 2002).

Data Analysis Overview

Our primary analyses were organized into two analytic Tiers. A priori, we specified Sample 1 as the primary sample and Sample 2 as the replication sample. In Tier 1, we used confirmatory factor analysis (CFA) to determine the best representation of children's hyperactivity across the experimental conditions (measurement models) as well as across parent and teacher hyperactivity ratings. Because the working memory set size conditions were administered on separate testing days for Sample 1 as described above, there were 4 indicators each for actigraphy during the phonological and visuospatial working memory conditions (1 indicator each from memory loads 3–6). In contrast, the set sizes in Sample 2 were presented in randomized/unpredictable order within a single task; thus, in Sample 2 actigraphy during the phonological and visuospatial working memory tasks were each modeled using single-indicator latents. The parent and teacher models were identical across studies, and used the four DSM-5-TR hyperactivity items identified in the Kofler et al. (2020) bifactor study as uniquely reflecting excess motor movement (as opposed to verbal intrusion/impulsivity).

In Tier 2, we used latent path modeling to correlate the objective actigraph latents with the subjective parent and teacher latents to test the extent to which clinic-based measurement of hyperactivity demonstrates similar levels of ecological validity as cross-setting (school vs. home) ratings. We then constrained pathways between/across actigraphy/ratings to examine whether correlations were significantly higher or lower for cross-domain (actigraphy/ratings) relative to within-domain associations (parent/teacher ratings, actigraphy across working memory/baseline conditions). In other words, we asked whether clinic-based measurement of hyperactivity was comparable to parent ratings of hyperactivity at home in terms of their ecological validity for predicting hyperactivity at school (and vice versa). Finally, sensitivity analyses were added to explore the impact of our a priori decision not to control for age and sex in the primary models.

For all confirmatory models, absolute and relative fit were tested. Adequate model fit was defined a priori as CFI and TFI ≥ .90, and RMSEA ≤ .10. Chi-squared difference tests ($\Delta\chi^2$) were used to compare nested models for both initial measurement model selection and the final structural model's pathway constraints.

All items showed the expected range of scores and were screened for normality. All variables of interest had acceptable skewness < |3| and kurtosis values for SEM (kurtosis

< |10|; Brown, 2006). Standardized residuals were inspected for magnitude (all positive and <1, indicating no evidence of localized ill fit). Missing data rates were low (Sample 1: 8.74%, Sample 2: 2.95% total), were missing completely at random (Little's MCAR: both p>.10), and were accounted for using full information maximum likelihood. Maximum likelihood estimation with robust standard errors (MLR) and delta scaling was specified a priori to handle any non-normality and non-independence of the data (Kline, 2016; Muthén & Muthén, 2017). Directionality of parameter estimates were inspected.

Power Analyses

A series of Monte Carlo simulations were run using Mplus 8.7 (Muthén & Muthén, 2002, 2021) to estimate the power of our proposed path model for each sample for detecting significant relations among hyperactivity variables measured by actigraphy and informant-rating scales, 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 and expected residual variances for observed variables informed by published studies using hyperactivity assessed using actigraphy and informant-rating (e.g., Kofler et al., 2020; Soto, Black, & Kofler, 2024) were imputed to delineate the proposed models. Results indicated that both sample models (sample 1 N=88 and sample 2 N=184) are powered to detect correlations of approximately r=.34. The similar power across samples occurred because the smaller Sample 1 included additional indicators, which improved latent variable reliability, thereby stabilizing structural correlations and increasing power. These power estimates are similar or lower than published correlations between actigraphy and symptom ratings in ADHD (r=.32-.58; Rapport, Kofler, & Himmerich, 2006) and other clinical populations (e.g., autism spectrum disorder, mood disorders; Kaufmann et al., 2021; Nadja et al., 2011; Bangerter et al., et al., 2020, respectively). Thus, the study is sufficiently powered to address our primary confirmatory model aims.

Results

Preliminary Analyses

All parent and teacher ADHD rating scale scores were higher for the ADHD relative to Non-ADHD group in both samples as expected (Tables 1a and 1b). THS activity level was also higher for the ADHD group relative to the comparison group across all conditions in both sample 1 (d=0.64–1.36, all p<.009) and sample 2 (d=0.41–0.68; all p<.01). There was no significant evidence to indicate between-group differences in socioeconomic status (p>.29) or race/ethnicity (p>.11) in either sample. In sample 1, the comparison group demonstrated slightly higher IQ scores (104 vs. 110, p=.009; no differences in sample 2, p=.26), whereas in sample 2 the ADHD group was approximately 5 months younger than the Non-ADHD group on average (p=.04; no differences in sample 1, p=.22).

Tier 1: Measurement Models

Actigraph models. Following Friedman et al. (2022), we initially created separate latent variables for lower (memory loads 3 and 4) and higher (memory loads 5 and 6) working memory demands based on the hypothesis that the lower memory loads would more

closely match typical working memory demands in classroom settings (Kofler et al., 2008). However, these models produced inadmissible results (e.g., intercorrelations greater than 1.0) and were therefore abandoned. The primary actigraph model therefore included 3 factors for actigraphy during (a) phonological working memory testing, (b) visuospatial working memory testing, and (c) the baseline painting conditions. As shown in Table 2, this model showed adequate fit in Sample 1, with all indicators loading significantly onto their respective latent constructs. These findings were replicated in Sample 2 (all CFI/TLI ≥ .90, RMSEA < .10). Notably, however, the latent correlation between hyperactivity during phonological and visuospatial working memory testing was very high in Sample 1 (r=.93) despite being only moderate in Sample 2 (r=.61). We therefore explored an alternative model that included a single working memory latent construct comprised of actigraphy across all phonological and visuospatial working memory task conditions.

Informant ratings models. As shown in Table 2, the model with separate parent and teacher hyperactivity ratings showed adequate fit in Sample 1, with all indicators loading significantly onto their respective latent constructs. These findings were replicated in Sample 2 (all CFI/TLI ≥ .90, RMSEA < .10).

Tier 2: Structural Models

Primary structural model. The primary Tier 2 structural model involved including the parent ratings, teacher ratings, baseline actigraphy, phonological working memory actigraphy, and visuospatial working memory actigraphy latents in the same model and allowing them to correlate. As shown in Table 3, the model showed adequate fit in Sample 1, with all indicators loading significantly onto their respective latent constructs. These findings were replicated in Sample 2 (all CFI/TLI ≥ .90, RMSEA < .10). All latent constructs were significantly intercorrelated (all p<.001). We then constrained pathways between/across actigraphy/ratings to examine whether correlations were significantly higher or lower for cross-domain (actigraphy/ratings) relative to within-domain (parent/teacher ratings, actigraphy across working memory/baseline conditions) associations. Constraining all pathways to be equal resulted is significantly worse model fit (p<.001). We therefore iteratively released pathway constraints until identifying the most parsimonious model that produced similar (not significantly different) fit relative to the fully unconstrained model.

As expected, the correlation between actigraphy during the two working memory tests was significantly larger than all other associations (r=.93), which we interpret as evidence for strong test-retest reliability. Interestingly, in terms of within-domain associations, the parent/teacher, baseline/visuospatial actigraphy, and baseline/phonological actigraphy ratings were statistically indistinguishable ($\Delta\chi^2$ [7] = 11.61, p=.11) from each other (both r=.57), suggesting that actigraphy across high and low cognitive demands shows the same (or better, when correlating actigraphy during the different types of working memory tests) level of agreement as parent and teacher hyperactivity ratings. In terms of cross-domain associations, the association between parent and teacher ratings was statistically indistinguishable from these ratings' associations with actigraphy during visuospatial working memory testing (r=.57). Stated differently, actigraphy during clinic-based visuospatial working memory testing predicts real-world behavior at home and school just as well as parent perceptions of

hyperactivity at home predict teacher perceptions of hyperactivity at school (and vice versa). The associations between parent/teacher ratings and actigraphy during the phonological working memory tasks and baseline conditions was somewhat lower, but remained robust (r=.44).

These findings were largely replicated in Sample 2 (Table 3). The final partially constrained model revealed that within-domain associations among actigraph latents were consistently stronger than the relation between the parent and teacher rating latents (r=.61 vs. .35). In terms of cross-domain associations, the association between parent and teacher ratings was statistically indistinguishable ($\Delta\chi^2$ [5] = −7.41, p=.99) from their associations with actigraphy during all 3 clinic-based conditions (all r=.35), with the exception of smaller associations between teacher ratings and actigraphy during phonological working memory testing (r=.21). Taken together, these findings indicate that actigraphy during clinic-based testing is superior to, or equivalent with, parent and teacher hyperactivity ratings in terms of consistency across sessions and types of activities. In addition, actigraphy during visuospatial working memory testing showed the same level of ecological validity as parent and teacher ratings in terms of its ability to predict hyperactivity in real-world school and home settings, respectively. Actigraphy during phonological working memory and baseline conditions also uniquely predicted hyperactivity at home and school, albeit with a less consistent pattern in terms of whether these associations were as strong as the parent/teacher association (Figure 1).

Exploratory structural model. Next, as noted above, given the high correlation between actigraphy during phonological and visuospatial working memory testing in Sample 1, we also tested an alternative model that included a single working memory latent (comprised of the phonological and visuospatial indicators from the primary model). The analyses mirrored the steps described in the primary model; reporting is truncated for readability. In both samples, actigraphy during the baseline and working memory conditions showed stronger convergence (r=.69-.79) than parent ratings' association with teacher ratings (r=.38-.52). Similarly, actigraphy showed strong evidence of ecological validity, with the association between parent and teacher ratings being statistically indistinguishable from both informants' associations with both actigraph latents across both samples, with the exception of a smaller teacher ratings/phonological actigraphy association in Sample 1. As with the primary model findings, the strength of associations was overall smaller in Sample 2 relative to Sample 1, which likely reflects the increased clinical heterogeneity of Sample 2 as well as the increased quantity of indicators for each actigraph latent in Sample 1.

Sensitivity Analyses

Sensitivity analyses were conducted by repeating the Tier 2 analysis above, this time covarying age and sex. Fit remained adequate for both samples (CFI=.95-.96, TLI=.93-.94, both RMSEA=.06). Older age predicted less hyperactivity based on parent report in Sample 1 (β=−.20, p=.01) but not teacher ratings or actigraphy scores (p>.15), whereas the opposite pattern was observed in Sample 2: Older age predicted teacher ratings and all actigraph latents (β=−.27 to −.30, all p<.001) but not parent report (β=−.08, p=.32). Sex did not predict any of the hyperactivity latents in Sample 1 (all p>.47), likely due

to the small number of girls in that sample.[3] In contrast, in Sample 2 boys exhibited greater hyperactivity based on both informants and all three actigraph latents (β=.21-.28, all p<.001). Interestingly, however, the pattern and interpretation of all findings remained unchanged in both samples with age and sex covaried, such that the correlation coefficients did not differ between the primary and exploratory analyses by more than $|r|$ = .01 (Sample 1) to .05 (Sample 2).

Discussion

The current study was the first to use latent path modeling to examine the ecological validity of clinic-based actigraphy assessment relative to a criterion benchmark – the association between cross-informant hyperactivity ratings. Additional strengths of the study included deliberate probes for robustness/replicability via the use of two independent samples with somewhat different compositions (e.g., inclusion vs. exclusion of common co-occurring conditions), the clinically evaluated and carefully phenotyped samples, and the multi-informant, multi-method design. Overall, the findings provide strong evidence supporting the ecological validity of clinic-based actigraphy assessment – especially for actigraphy during visuospatial working memory testing. These findings replicated across two independent samples and showed robust associations with parent- and teacher-rated hyperactivity that were statistically indistinguishable from the association between these informants' perceptions of child hyperactive behavior. Actigraphy during phonological working memory testing and our baseline conditions also showed ecological validity evidence; however, in some cases, the magnitude of these associations was somewhat smaller than our criterion parent/teacher benchmark. Implications of these findings are discussed below.

Ecological validity refers to the extent to which a test or measure predicts real-world functioning (Andrade, 2018). Critically, however, it is often conflated with face validity (Suchy et al., 2024a, 2024b) – the extent to which a test resembles real-world situations – despite evidence that face validity likely does not enhance ecological validity (Ziemnik & Suchy, 2019). Indeed, a recent review noted that a substantial proportion of studies claiming 'ecological validity' of clinic-based measures were based on qualitative evaluation rather than quantitative comparison of the measures' relation with relevant real-world outcomes (Suchy et al., 2024). To that end, the current study quantitatively compared cross-method associations between actigraphy and 'real world' behavior as perceived by multiple informants (parents, teachers). Broadly, we found that actigraphy during working memory testing – particularly visuospatial working memory testing – showed similar associations with both parent and teacher ratings as those informant ratings showed with each other. We also found that actigraphy scores were generally superior to informant ratings in terms of test-retest/internal validity for predicting hyperactivity across settings or test type/testing days. The cross-method (r=.35-.57) and within-method (ratings: r=.35-.57; actigraphy within test type: r=.61-.93; actigraphy across test type: r=.35-.61) associations detected across samples in the current study are broadly consistent with prior evidence for (a) relations

[3]We also repeated these sensitivity analyses in Sample 1 covarying just age, as well as repeating the analyses with the girls (n=5) excluded (with and without covarying age). The pattern and interpretation of the findings was unchanged in all cases.

between clinic-based actigraphy informant ratings from other settings (r=.32-.58; Rapport et al., 2008); (b) relations between parent and teacher ADHD symptom ratings (r=.10-.50; Saffer et al., 2021; Takeda et al., 2020); and (c) reliability of actigraphy ≥.90 (Tryon et al., 1991).

Thus, the most parsimonious conclusion is that clinic-based actigraphy provides robust test-retest reliability (r=.61-.93 across different working memory tests administered on different testing days) and concurrent validity (r=.35-.61 across tasks with high vs. low cognitive demands). Central to this study's main aims, clinic-based actigraphy also provides equivalent (or not significantly inferior) external/ecological validity relative to parent ratings for predicting school-based (teacher rated) hyperactivity. Of course, given the correlational nature of these data, it is equally correct to say that actigraphy is just as ecologically valid for predicting hyperactivity at home relative to school-based teacher ratings, as it is to say that parent ratings are just as ecologically valid for predicting teacher ratings, and so on. To that end, we contend that the evidence base at this time indicates that objective measurement of hyperactivity in the clinic (i.e., via actigraphy) provides an ecologically valid assessment of children's hyperactivity at home and at school, to the extent that actigraphy is obtained during tasks that sufficiently challenge these children executive functions (Soto et al., 2024). Indeed, recent clinical trials suggest that clinic-based actigraphy maybe useful for assessing the comparative efficacy of psychosocial treatments for ADHD (Kofler et al., 2018; cf. Kofler et al., 2020).

Nonetheless, we acknowledge that it is premature to advocate for the use of actigraphy in routine clinical practice given that nationally representative norms (during criterion tests) are not currently available. Thus, the utility of actigraphy in routine clinical practice is currently limited to clinics that have collected a sufficient local normative sample. However, we predict that this will change over the next decade given evidence that both criterion working memory tests (for reviews, see Kofler et al., 2016; Soto et al., 2024) both tend to show larger ADHD/Non-ADHD between-group effect sizes than continuous performance tests and other clinic-based measures (Huang-Pollock et al. 2013; Kofler et al., 2024; Rapport et al., 2001). Given these findings and the corresponding reliability/validity evidence as described above, it seems highly likely that the combination of performance and actigraphy during criterion working memory tests may provide improved diagnostic utility relative to existing clinic-based tests – particularly for differentiating ADHD from other conditions with overlapping symptoms (e.g., concentration problems/restlessness in anxiety and depression; Cibrian et al., 2025).

Indeed, at least one commercially available test uses a combination of test performance and (video camera detected) movement to predict ADHD diagnostic status (Reh et al., 2015); however, in that case movement is assessed during a monotonous task similar to continuous performance/choice response tests that generally do not differentially evoke hyperactivity in children with vs. without ADHD (Soto et al., 2024) – a likely artifact of the near-ubiquitous but questionable 'jingle fallacy' assumption that the behavioral attention deficits in ADHD are due to a neurocognitive attention deficit (e.g., Kofler et al., 2024). Going forward, the current findings underscore the need to evaluate hyperactive behavior while children are engaged in cognitively challenging tasks that place sufficient demands on working

memory (Soto et al., 2024) or other executive processes empirically linked with hyperactive (and inattentive) behavior via experimental, longitudinal, and clinical trial evidence (e.g., Hartanto et al., 2015; Karalunas et al., 2017; Kofler et al., 2008, 2018, 2020; Rapport et al., 2009; Sarver et al., 2015).

Limitations

The current study has several strengths, including the clinically evaluated and carefully phenotyped samples, latent path modeling, and replication across two independent samples. At the same time, the following limitations should be considered. First, the independent samples differed in several ways, including inclusion criteria, certain demographics, and the extent to which those demographics predicted hyperactivity. Curiously, age was associated with different informant ratings (parent vs. teacher) and in different directions (younger vs. older age associated with greater activity level) across the samples. This may reflect the somewhat different mean ages across the samples (9.4 vs. 10.2 years) combined with the developmentally curvilinear relation between activity level and positive outcomes (Rapport et al., 2006), differences in inclusion/exclusion of co-occurring clinical conditions, and/or differences in the ratio of boys to girls across the samples. Alternatively, given that age and sex did not significantly change any of the primary model results, these demographic findings may just reflect chance variation around true effect sizes of 0.0 given the relatively narrow age range (8–12) in both samples. Relatedly, despite the study's relatively large sample size, participants that identified as White or reported higher maternal education (i.e., bachelor's degree or higher) were disproportionally represented relative to other racial identities and education levels. Therefore, generalizability to other populations – including historically underrepresented racial minority groups and those with mothers of lower education levels – may be limited.

Next, the results differed somewhat based on whether actigraphy was assessed during visuospatial vs. phonological working memory testing. Specifically, across both samples actigraphy during visuospatial working memory testing reliably demonstrated associations with both parent and teacher ratings that were indistinguishable from those ratings' associations with each other. In contrast, this result was found in only one of four comparisons for actigraphy during phonological working memory testing. We suspect that this occurred due to small but likely meaningful differences in task demands. That is, the visuospatial tasks required children to respond using a modified keypad, versus auditory responses for the phonological tasks. Thus, children needed to stay closer to the computer during the visuospatial tasks relative to the more 'free range' phonological tasks – with the latter allowing more variability in movement. Inspection of Tables 1a-b provide tentative support for this hypothesis, with variability (SD) generally 40–50% higher during the phonological relative to visuospatial tasks.

Finally, we *a priori* set our benchmark for actigraphy's ecological validity to be the association between teacher and parent ratings of hyperactivity in each sample. This is admittedly a lofty comparison given that associations are typically inflated for mono-method (e.g., ratings/ratings) vs. cross-method (e.g., ratings/actigraphy) comparisons – i.e., due to mono-method bias (Podsakoff et al., 2003). Nonetheless, we reasoned that informant ratings

were the apt comparison given that they are considered to reflect behavior in real-world settings and are frequently used as the criterion for studies assessing ecological validity in the ADHD literature. Future work may benefit from comparison to more objective evaluations of real-world behavior (e.g., direct observations; Kofler et al., 2008; Rapport et al., 2009).

Practitioner Implications and Future Directions

Collectively, the current study demonstrated across two independent samples that clinic-based actigraphy – particularly when assessed during visuospatial working memory tasks – are likely comparable to parent and teacher hyperactivity ratings in terms of their ecological validity. At the same time, to our knowledge, nationally representative normative data are not available, which may currently limit the clinical utility of these findings. Notably, however, hospital-based and other practices that evaluate large numbers of children per year could quickly – and relatively inexpensively – build a local normative sample of actigraphy during neuropsychological testing that may provide incremental validity for differentiating ADHD from other clinical conditions (Faust et al., 2024). Looking ahead, we predict that the combination of test performance and actigraphy during tests with high cognitive demands (e.g., working memory) may hold promise for upgrading the science and technology of ADHD assessment despite the underwhelming performance of traditional neuropsychological tests that the DSM-5-TR explicitly notes are not useful for ADHD diagnosis (APA, 2022; Rapport et al., 2001).

Data, Materials and/or Code Availability:

Anonymized data/code and results output are available from the authors upon reasonable request.

Acknowledgements:

This work was supported in part by NIH grants (R01MH125893, R61MH127031; 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

Achenbach TM, & Rescorla LA (2001). Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families

Alderson RM, Rapport MD, Kasper LJ, Sarver DE, & Kofler MJ (2012). Hyperactivity in boys with attention deficit/hyperactivity disorder (ADHD): The association between deficient behavioral inhibition, attentional processes, and objectively measured activity. Child Neuropsychology, 18(5), 487–505. 10.1080/09297049.2011.631905

Ambulatory Monitoring, Inc. (2014a). Micro Motionlogger actigraph [Apparatus]. https://www.ambulatory-monitoring.com/

Ambulatory Monitoring, Inc. (2014b). Motionlogger MicroMini actigraph [Apparatus]. https://www.ambulatory-monitoring.com/

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). 10.1176/appi.books.9780890425787

Andrade C (2018) Internal, external, and ecological validity in research design, conduct, and evaluation. Indian Journal of Psychological Medicine, 40(5), 498–499. 10.4103/IJPSYM.IJPSYM_334_18

Bangerter A, Chatterjee M, Manyakov NV, Ness S, Lewin D, Skalkin A, Boice M, Goodwin MS, Dawson G, Hendren R, Leventhal B, Shic F, Esbensen A, & Pandina G (2020). Relationship between sleep and behavior in autism spectrum disorder: Exploring the impact of sleep variability. Frontiers in Neuroscience, 14, 211. 10.3389/fnins.2020.00211

Barkley RA (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. Journal of Abnormal Child Psychology, 19, 149–178. 10.1007/BF00909976

Barkley RA (2016). Recent longitudinal studies of childhood attention-deficit/hyperactivity disorder: Important themes and questions for further research. Journal of Abnormal Psychology, 125(2), 248–255. 10.1037/abn0000125

Barkley RA (2019). Neuropsychological testing is not useful in the diagnosis of ADHD: Stop it (or prove it)! The ADHD Report, 27(2), 1–8. 10.1521/adhd.2019.27.2.1

Brown TA (2006). Confirmatory factor analysis for applied research. New York, NY: The Guilford Press.

Chang TM, Wu RC, Yang RC, Chiang CT, Chiu YH, Ouyang CS, Wang YM, Wu MH, Lin GC, & Lin LC (2023). Objective diagnosis of ADHD through movement analysis by using a smart chair with piezoelectric material. Pediatrics and Neonatology, 64(1), 46–52. 10.1016/j.pedneo.2022.06.007

Cibrian E, Padron C, Bruce A, Cho S, Tatsuki Patron, B., Geren MMO, Irwin Harper LN, & Kofler MJ (2025, January). Examining objectively-measured motor activity among children with and without ADHD and GAD. [Poster presentation]. The American Professional Society of ADHD and Related Disorders (APSARD), Annual Conference, San Diego, CA, United States.

Cirino PT, Chin CE, Sevcik RA, Wolf M, Lovett M, & Morris RD (2002). Measuring socioeconomic status: reliability and preliminary validity for different approaches. Assessment, 9(2), 145–155. 10.1177/10791102009002005

Dane AV, Schachar RJ, & Tannock R (2000). Does actigraphy differentiate ADHD subtypes in a clinical research setting? Journal of the American Academy of Child and Adolescent Psychiatry, 39(6),752–760. 10.1097/00004583-200006000-00014

Danielson ML, Claussen AH, Bitsko RH, Katz SM, Newsome K, Blumberg SJ, Kogan MD, & Ghandour R (2024). ADHD Prevalence among US children and adolescents in 2022: Diagnosis, severity, co-occurring disorders, and treatment. Journal of Clinical Child & Adolescent Psychology, 53(3), 343–360. 10.1080/15374416.2024.2335625

De Crescenzo F, Licchelli S, Ciabattini M, Menghini D, Armando M, Alfieri P, Mazzone L, Pontrelli G, Livadiotti S, Foti F, Quested D, & Vicari S (2016). The use of actigraphy in the monitoring of sleep and activity in ADHD: A meta-analysis. Sleep Medicine Reviews, 26, 9–20. 10.1016/j.smrv.2015.04.002

Dekkers TJ, Rapport MD, Calub CA, Eckrich SJ, & Irurita C (2021). ADHD and hyperactivity: The influence of cognitive processing demands on gross motor activity level in children. Child Neuropsychology, 27(1), 63–82. 10.1080/09297049.2020.1793924

DuPaul GJ, Kern L, Caskie GI, & Volpe RJ (2015). Early intervention for young children with attention deficit hyperactivity disorder: Prediction of academic and behavioral outcomes. School Psychology Review, 44(1), 3–20. 10.17105/SPR44-1.3-20

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

DuPaul GJ, & Stoner G (2014). ADHD in the schools: Assessment and intervention strategies. New York, NY: Guilford Publications.

Fabiano GA, Schatz NK, Aloe AM, Pelham WE, Smyth AC, Zhao X, Merrill BM, Macphee F, Ramos M, Hong N, Altszuler A, Ward L, Rodgers DB, Liu Z, Karatoprak Ersen R, & Coxe S (2021). Comprehensive meta-analysis of attention-deficit/hyperactivity disorder psychosocial treatments investigated within between group studies. Review of Educational Research, 91(5), 718–760. 10.3102/00346543211025092

Faust D, Arkes HR, & Gaudet CE (2024). Applying decision research to improve clinical outcomes, psychological assessment, and clinical prediction. New York, NY: Oxford University Press.

Friedman LM, Rapport MD, & Fabrikant-Abzug G (2022). Consistently inconsistent working memory performance among children with ADHD: Evidence of response accuracy variability (RAV). Journal of Psychopathology and Behavioral Assessment, 44(3), 787–799. 10.1007/s10862-022-09967-7

Gadow KD, Sprafkin J, Salisbury H, Schneider J, and Loney J (2004). Further validity evidence for the teacher version of the Child Symptom lnventory-4. School Psychology Quarterly, 19(1), 50–71. 10.1521/scpq.19.1.50.29408

Gomez R, & Sanson AV (1994). Effects of experimenter and mother presence on the attentional performance and activity of hyperactive boys. Journal of Abnormal Child Psychology, 22, 517–529. 10.1007/BF02168935

Halperin JM, Newcorn JH, Matier K, Sharma V, McKay KE, & Schwartz S (1993). Discriminant validity of attention deficit hyperactivity disorder. Journal of the American Academic of Child and Adolescent Psychiatry, 32(5), 1038–1043. 10.1097/00004583-199309000-00024

Hartanto TA, Krafft CE, Iosif AM, & Schweitzer JB (2015). A trial-by-trial analysis reveals more intense physical activity is associated with better cognitive control performance in attention-deficit/hyperactivity disorder. Child Neuropsychology, 22(5), 618–626. 10.1080/09297049.2015.1044511

Huang-Pollock CL, Nigg JT, & Carr TH (2005). Deficient attention is hard to find: Applying the perceptual load model of selective attention to attention deficit hyperactivity disorder subtypes. Journal of Child Psychology and Psychiatry, 46(11), 1211–1218. 10.1111/j.1469-7610.2005.00410.x

Hudec KL, Alderson RM, Patros CH, Lea SE, Tarle SJ, & Kasper LJ (2015). Hyperactivity in boys with attention-deficit/ hyperactivity disorder (ADHD): The role of executive and non-executive functions. Research in Developmental Disabilities, 45-46, 103–109. 10.1016/J.RIDD.2015.07.012

JASP Team (2024). JASP (Version 0.19.1) [Computer software]. https://jasp-stats.org/

Karalunas SL, Gustafsson HC, Dieckmann NF, Tipsord J, Mitchell SH, & Nigg JT (2017). Heterogeneity in development of aspects of working memory predicts longitudinal attention deficit hyperactivity disorder symptom change. Journal of Abnormal Psychology, 126(6), 774–792. 10.1037/abn0000292

Karalunas SL, Huang-Pollock CL, & Nigg JT (2013). Is reaction time variability in ADHD mainly at low frequencies? Journal of Child Psychology and Psychiatry, 54(5), 536–544. 10.1111/jcpp.12028

Kaufman J, Birmaher B, Brent D, Rao UMA, Flynn C, Moreci P, Williamson D, & Ryan N (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. Journal of the American Academy of Child & Adolescent Psychiatry, 36(7), 980–988. 10.1097/00004583-199707000-00021

Kaufman AS, & Kaufman NL (2014). Technical & interpretive manual: Kaufman Test of Educational Achievement (3rd ed.). Bloomington, MN: NCS Pearson.

Kaufmann CN, Lee EE, Wing D, Sutherland AN, Christensen C, Ancoli-Israel S, Depp CA, Yoon HK, Soontornniyomkij B, & Eyler LT (2021). Correlates of poor sleep based upon wrist actigraphy data in bipolar disorder. Journal of Psychiatric Research, 141, 385–389. 10.1016/j.jpsychires.2021.06.038

Kline RB (2016). Principles and Practice of Structural Equation Modeling (4th ed.). New York NY: The Guilford Press.

Kofler MJ, Groves NB, Chan ES, Marsh CL, Cole AM, Gaye F, Cibrian E, Tatsuki MO, & Singh LJ (2024). Working memory and inhibitory control deficits in children with ADHD: An experimental evaluation of competing model predictions. Frontiers in Psychiatry, 15. 10.3389/fpsyt.2024.1277583

Kofler MJ, Groves NB, Singh LJ, Soto EF, Chan ESM, Irwin LN, & Miller CE (2020). Rethinking hyperactivity in pediatric ADHD: Preliminary evidence for a reconceptualization of hyperactivity/impulsivity from the perspective of informant perceptual processes. Psychological Assessment, 32(8), 752–767. 10.1037/pas0000856

Kofler MJ, Raiker JS, Sarver DE, Wells EL, & Soto EF (2016). Is hyperactivity ubiquitous in ADHD or dependent on environmental demands? Evidence from meta-analysis. Clinical Psychology Review, 46, 12–24. 10.1016/j.cpr.2016.04.004

Kofler MJ, Rapport MD, & Alderson RM (2008). Quantifying ADHD classroom inattentiveness, its moderators, and variability: A meta-analytic review. Journal of Child Psychology and Psychiatry, 49(1), 59–69. 10.1111/j.1469-7610.2007.01809.x

Kofler MJ, Spiegel JA, Austin KE, Irwin LN, Soto EF, & Sarver DE (2018). Are episodic buffer processes intact in ADHD? Experimental evidence and linkage with hyperactive behavior. Journal of Abnormal Child Psychology, 46, 1171–1185. 10.1007/s10802-017-0346-x

Kofler MJ, Spiegel JA, Soto EF, Irwin LN, Wells EL, & Austin KE (2019). Do working memory deficits underlie reading problems in attention-deficit/hyperactivity disorder (ADHD)? Journal of Abnormal Child Psychology, 47, 433–446. 10.1007/s10802-018-0447-1

Lahey BB, Pelham WE, Loney J, Lee SS, & Willcutt E (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. Archives of General Psychiatry, 62(8), 896–902. 10.1001/archpsyc.62.8.896

Lee WH, Kim JI, Kwon AM, Cha JH, Yim D, Lim YH, Cho SH, Cho SH, & Park HK (2021). Quantified assessment of hyperactivity in ADHD youth using IR-UWB radar. Scientific Reports, 11(1), 9604. 10.1038/s41598-021-89024-7

Murillo LG, Cortese S, Anderson D, Di Martino A, & Castellanos FX (2015). Locomotor activity measures in the diagnosis of attention deficit hyperactivity disorder: Meta-analyses and new findings. Journal of Neuroscience Methods, 252, 14–26. 10.1016/j.jneumeth.2015.03.001

Muthén BO, & Muthén LK (2017). Mplus. In van der Linden WJ (Ed.), Handbook of item response theory: Volume 3: Applications (1st ed., pp. 507–518). Boca Raton, FL: Chapman and Hall/CRC.

Muthén LK, & Muthén BO (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural equation modeling, 9(4), 599–620. 10.1207/S15328007SEM0904_8

Muthén LK, & Muthén BO (2021). Mplus (Version 8.7) [Computer software]. https://www.statmodel.com/

Noldus Information Technology (2014). The Observer XT [Computer software]. https://www.noldus.com/observer-xt-human

Orban SA, Rapport MD, Friedman LM, Eckrich SJ, & Kofler MJ (2018). Inattentive behavior in boys with ADHD during classroom instruction: The mediating role of working memory processes. Journal of Abnormal Child Psychology, 46, 713–727. 10.1007/s10802-017-0338-x

Patros C, Alderson RM, Hudec KL, Tarle SJ, & Lea SE (2017). Hyperactivity in boys with attention-deficit/hyperactivity disorder: The influence of underlying visuospatial working memory and self-control processes. Journal of Experimental Child Psychology, 154, 1–12. 10.1016/j.jecp.2016.09.008

Plomin R, & Foch TT (1981). Hyperactivity and pediatrician diagnoses, parental ratings, specific cognitive abilities, and laboratory measures. Journal of Abnormal Child Psychology, 9, 55–64. 10.1007/BF00917857

Podsakoff PM, MacKenzie SB, Lee JY, & Podsakoff NP (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879–903. 10.1037/0021-9010.88.5.879

Polanczyk GV, Willcutt EG, Salum GA, Kieling C, & Rohde LA (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. International Journal of Epidemiology, 43(2), 434–442. 10.1093/ije/dyt261

Porrino LJ, Rapoport JL, Behar D, Sceery W, Ismond DR, & Bunney WE (1983). A naturalistic assessment of the motor activity of hyperactive boys: I. Comparison with normal controls. Archives of General Psychiatry, 40(6), 681–687. 10.1001/archpsyc.1983.04390010091012

Rapport MD, Alderson RM, Kofler MJ, Sarver DE, Bolden J, & Sims V (2008). Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): The contribution of central executive and subsystem processes. Journal of Abnormal Child Psychology, 36, 825–837. 10.1007/s10802-008-9215-y

Rapport MD, Bolden J, Kofler MJ, Sarver DE, Raiker JS, & Alderson RM (2009). Hyperactivity in boys with attention-deficit/hyperactivity disorder (ADHD): a ubiquitous core symptom or

manifestation of working memory deficits? Journal of Abnormal Child Psychology, 37, 521–534. 10.1007/s10802-008-9287-8

Rapport MD, Chung KM, Shore G, & Isaacs P (2001). A conceptual model of child psychopathology: Implications for understanding attention deficit hyperactivity disorder and treatment efficacy. Journal of Clinical Child & Adolescent Psychology, 30(1), 48–58. 10.1207/S15374424JCCP3001_6

Rapport MD, Kofler MJ, Alderson RM, Timko TM, & DuPaul GJ (2009). Variability of attention processes in ADHD: Observations from the classroom. Journal of Attention Disorders, 12(6), 563–573. 10.1177/1087054708322990

Rapport MD, Kofler M, & Himmerich C (2006). Assessing activity level in children. In Hersen M (Ed.), Clinician's Handbook of Child Behavioral Assessment (pp. 125–158). Burlington, MA: Elsevier Academic Press.

Razavi N, Horn H, Koschorke P, Hügli S, Höfle O, Müller T, Strik W, & Walther S (2011). Measuring motor activity in major depression: The association between the Hamilton Depression Rating Scale and actigraphy. Psychiatry Research, 190(2–3), 212–216. 10.1016/j.psychres.2011.05.028

Reh V, Schmidt M, Lam L, Schimmelmann BG, Hebebrand J, Rief W, & Christiansen H (2015). Behavioral assessment of core ADHD symptoms using the Qb Test. Journal of Attention Disorders, 19(12), 1034–1045.

Reynolds CR, & Kamphaus RW (2004). Behavior assessment system for children (2nd ed.). Bloomington, MN: Pearson.

Saffer BY, Mikami AY, Qi H, Owens JS, & Normand S (2021). Factors related to agreement between parent and teacher ratings of children's ADHD symptoms: An exploratory study using polynomial regression analyses. Journal of Psychopathology and Behavioral Assessment, 43, 793–807. 10.1007/s10862-021-09892-1

Sarver DE, Rapport MD, Kofler MJ, Raiker JS, & Friedman LM (2015). Hyperactivity in attention-deficit/hyperactivity disorder (ADHD): Impairing deficit or compensatory behavior? Journal of Abnormal Child Psychology, 43(7), 1219–1232. 10.1007/s10802-015-0011-1

Sleator EK, & Ullmann RK (1981). Can the physician diagnose hyperactivity in the office? Pediatrics, 67(1), 13–17. 10.1542/peds.67.1.13

Soto EF, Black K, & Kofler MJ (2024). Is hyperactivity in children with attention deficit/hyperactivity disorder (ADHD) a functional response to demands on specific executive functions or cognitive demands in general? Neuropsychology, 38(8), 699–713. 10.1037/neu0000975

Spiegel JA, Goodrich JM, Morris BM, Osborne CM, & Lonigan CJ (2021). Relations between executive functions and academic outcomes in elementary school children: A meta-analysis. Psychological Bulletin, 147(4), 329–351. 10.1037/bul0000322

Suchy Y, DesRuisseaux LA, Mora MG, Brothers SL, & Niermeyer MA (2024). Conceptualization of the term "ecological validity" in neuropsychological research on executive function assessment: A systematic review and call to action. Journal of the International Neuropsychological Society, 30(5), 499–522. 10.1017/S1355617723000735

Suchy Y, Mora MG, DesRuisseaux LA, Niermeyer MA, & Brothers SL (2024). Pitfalls in research on ecological validity of novel executive function tests: A systematic review and a call to action. Psychological Assessment. 36(4), 243–261. 10.1037/pas0001297

Takeda T, Nissley-Tsiopinis J, Nanda S, & Eiraldi R (2020). Factors associated with discrepancy in parent–teacher reporting of symptoms of ADHD in a large clinic-referred sample of children. Journal of Attention Disorders, 24(11), 1605–1615. 10.1177/1087054716652476

Tryon WW, Pinto LP, & Morrison DF (1991). Reliability assessment of pedometer activity measurements. Journal of Psychopathology and Behavioral Assessment, 13, 27–44. 10.1007/BF00960737

Valo S, & Tannock R (2010). Diagnostic instability of DSM–IV ADHD subtypes: Effects of informant source, instrumentation, and methods for combining symptom reports. Journal of clinical child & adolescent psychology, 39(6), 749–760. 10.1080/15374416.2010.517172

Wechsler D (2014). Wechsler Intelligence Scale for Children (5th ed.). Bloomington, MN: Pearson.

Wells EL, Kofler MJ, Soto EF, Schaefer HS, & Sarver DE (2018). Assessing working memory in children with ADHD: Minor administration and scoring changes may improve digit span

backward's construct validity. Research in developmental disabilities, 72, 166–178. 10.1016/j.ridd.2017.10.024

Willcutt EG, Nigg JT, Pennington BF, Solanto MV, Rohde LA, Tannock R, Loo SK, Carlson CL, McBurnett K & Lahey BB (2012). Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. Journal of abnormal psychology, 121(4), 991–1010. 10.1037/a0027347

Ziemnik RE, & Suchy Y (2019). Ecological validity of performance-based measures of executive functions: Is face validity necessary for prediction of daily functioning? Psychological Assessment, 31(11), 1307–1318. 10.1037/pas0000751

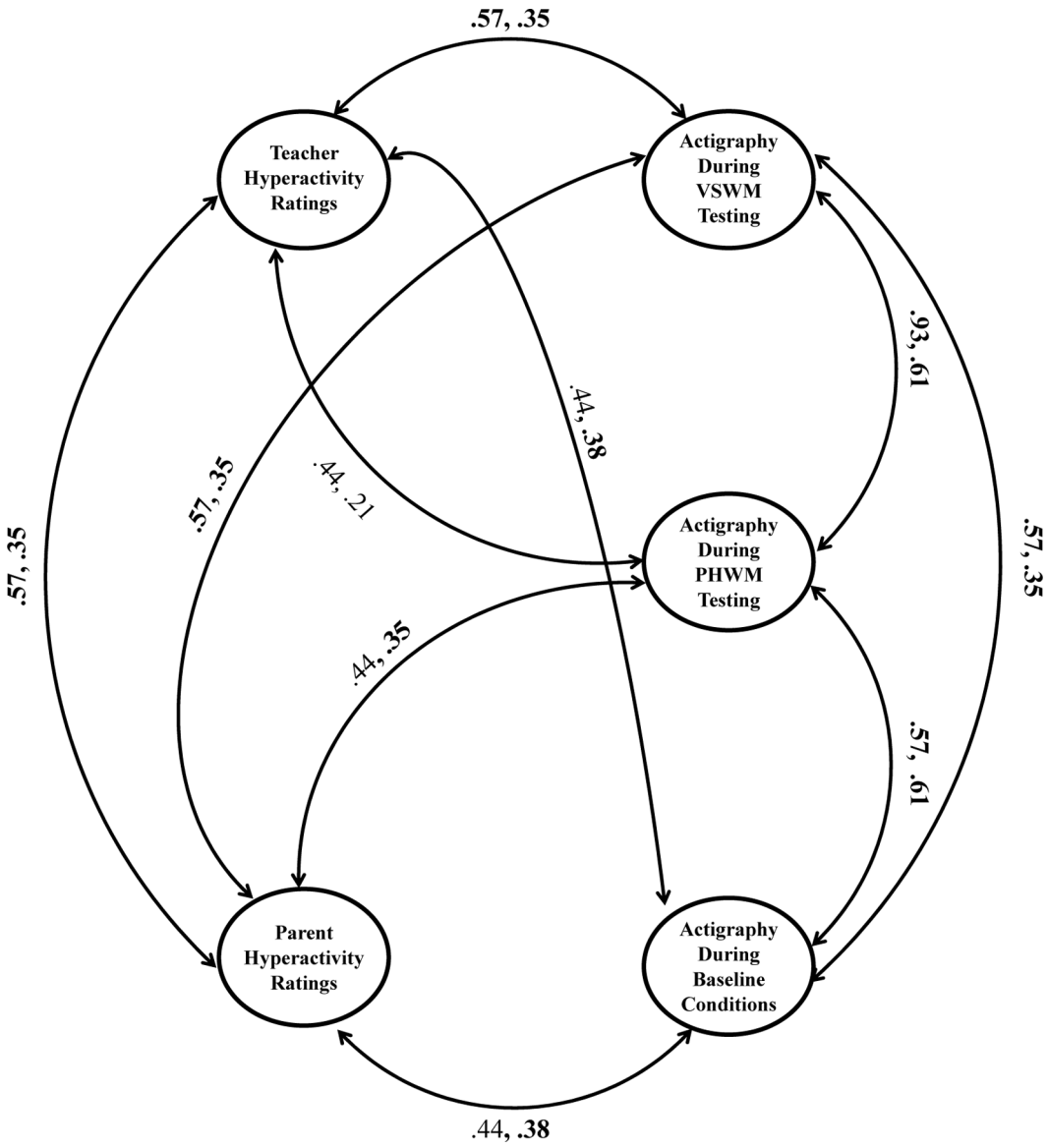

Figure 1.

Standardized structural model with parent and teacher-reported hyperactivity, as well as actigraphy during baseline, phonological working memory, and visuospatial working memory testing. The preferred constrained models are shown. Correlations are reported as Sample 1, Sample 2. Bolded pathways reflect relations that are equivalent to or larger than the criterion parent-teacher association benchmark for that Sample. All associations are significant at $p < .007$.

Table 1a.

Demographic characteristics (Sample 1).

	ADHD		Non-ADHD			
Variable	***M***	***SD***	***M***	***SD***	***Cohen's d***	***p***
N (Girls/Boys)	3/47		2/36		--	.88
Race/Ethnicity (A/B/H/M/W)	2/1/10/4/32		0/3/6/2/27		--	.44
FSIQ	104.06	11.77	110.50	10.44	0.57	.009
Age	9.40	1.26	9.94	1.27	0.22	.22
SES	47.68	12.41	50.33	15.10	0.19	.37
Attention Problems (CSI T-scores)						
Parent	75.33	8.57	51.50	13.03	2.21	<.001
Teacher	67.23	8.02	48.66	8.63	2.24	<.001
Hyperactivity/Impulsivity (CSI T-scores)						
Parent	71.77	15.52	46.42	9.23	1.93	<.001
Teacher	65.13	11.94	47.31	6.39	1.78	<.001
Actigraph Data (THS)						
Paint-Beginning	17111.69	6420.15	11198.15	4436.24	1.06	<.001
Paint-End	18786.22	6852.69	14652.46	6120.19	0.64	.009
VSWM Set Size 3	39228.91	10821.80	25420.06	9437.97	1.36	<.001
VSWM Set Size 4	36058.81	16295.09	22945.19	9331.76	0.98	<.001
VSWM Set Size 5	34572.51	10847.38	25423.99	9757.05	0.89	<.001
VSWM Set Size 6	36551.53	10860.46	23598.17	9090.45	1.29	<.001
PHWM Set Size 3	40444.50	14173.73	28899.74	15044.56	0.79	<.001
PHWM Set Size 4	43120.89	14443.70	28484.94	11158.64	1.13	<.001
PHWM Set Size 5	43711.02	10733.04	30457.70	14235.70	1.05	<.001
PHWM Set Size 6	43955.15	12475.20	31315.52	11370.35	1.06	<.001

Note: Ethnicity: A = Asian, B = Black, H = Hispanic, M = Multiracial, W = White Non-Hispanic. VCI = Verbal Comprehension Index. CSI = Child Symptom Inventory-IV. THS = Total Hyperactivity Score. PHWM = Phonological working memory. VSWM = Visuospatial working memory.

Table 1b.

Demographic characteristics (Sample 2).

	ADHD		Non-ADHD			
Variable	***M***	***SD***	***M***	***SD***	***Cohen's d***	***p***
N (Girls/Boys)	36/83		25/40		--	.33
Race/Ethnicity (A/B/H/M/W)	0/18/6/8/87		1/6/7/9/42		--	.11
VCI	103.81	14.29	106.13	11.38	0.17	.26
Age	10.24	1.48	10.70	1.49	0.31	.04
SES	47.83	11.15	49.71	11.43	0.17	.29
Attention Problems (BASC T-scores)						
Parent	67.45	8.86	59.68	12.76	0.75	<.001
Teacher	65.03	9.87	55.65	14.42	0.80	<.001
Hyperactivity/Impulsivity (BASC T-scores)						
Parent	68.19	13.97	56.92	14.10	0.80	<.001
Teacher	63.46	16.74	52.57	13.79	0.69	<.001
Actigraph Data (THS)						
Paint-Beginning	37.16	29.95	25.61	24.53	0.41	.01
Paint-End	59.06	47.77	38.30	28.46	0.49	<.001
Visuospatial Working Memory	172.41	105.90	106.58	78.25	0.68	<.001
Phonological Working Memory	246.41	151.81	157.30	101.25	0.66	<.001

Note: Ethnicity: A = Asian, B = Black, H = Hispanic, M = Multiracial, W = White Non-Hispanic. VCI = Verbal Comprehension Index. BASC = Behavior Assessment System for Children. THS = Total Hyperactivity Score.

Table 2.

Measurement Models

Model	CFI	TLI	RMSEA (90%CI)	SRMR	AIC	BIC	Factor loadings (all p<.001)	Latent correlations (all p<.001)
Sample 1 Actigraphs								
Separate PH, VS, and BL factors	.92	.90	.095 (.06-.13)	.07	2018	2107	BL: .65-.77 PH: .61-.89 VS: .65-.84	BL-PH: r=.60 BL-VS: r=.70 PH-VS: r=.93
Separate WM and BL factors	.91	.90	.098 (.06-.13)	.07	2020	2103	BL: .66-.75 WM: .62-.84	BL-WM: r=.67
Sample 2 Actigraphs								
Separate PH, VS, and BL factors	.99	.95	.001 (.00-.18)	.006	1660	1700	BL: .71-.79	BL-PH: r=.62 BL-VS: r=.59 PH-VS: r=.61
Separate WM and BL factors	.99	.99	.001 (.00-.18)	.006	1660	1700	BL: .71-.79 WM: .76-.80	BL-WM: r=.77
Sample 1 Ratings								
Separate parent & teacher ratings	.94	.92	.05 (.01-.23)	.04	1384	1446	Parent: .89-.96 Teacher: .61-.87	P-T: r=.66
Sample 2 Ratings								
Separate parent & teacher ratings	.97	.95	.08 (.05-.12)	.05	3044	3122	Parent: .75-.80 Teacher: .68-.90	P-T: r=.51

Note. AIC = Akaike information criteria, BIC = Bayesian information criteria, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, TLI = Tucker-Lewis index. BL = Baseline painting condition. P = Parent. PH = Phonological working memory. T = Teacher. VS = Visuospatial working memory. WM = working memory.

Table 3.

Structural Models

Model	CFI	TLI	RMSEA (90%CI)	$\Delta\chi^2$ [df]	Factor Loadings (all p<.001)	Latent correlations (all p<.007)
Sample 1						
Separate PH, VS, and BL factors						
Unconstrained	.94	.93	.07 (.04-.09)	--	All: .61-.95	P-T (r=.66), BL-VS (r=.70), BL-PH (r=.59), PH-VS (r=.93), T-BL (r=.37), T-PH (r=.35), T-VS (r=.39), P-VS (r=.53), P-BL (r=.47), P-PH (r=.39)
Constrained	.93	.92	.07 (.05-.09)	11.61 [7], p=.11	All: .61-.95	PH-VS (r=.99) > P-T = T-VS = P-VS = BL-VS = BL-PH (r=.57) > T-BL = T-PH = P-BL = P-PH (r=.44)
Separate WM and BL factors						
Unconstrained	.93	.92	.07 (.05-.09)	--	All: .61-.95	P-T (r=.66), BL-WM (r=.67), T-BL (r=.36), T-WM (r=.38), P-BL (r=.46), P-WM (r=.47)
Constrained	.93	.92	.07(.05-.09)	7.57 [3], p=.07	All: .61-.95	BL-WM (r=.69) > P-T = T-WM = P-WM (r=.52) > T-BL = P-BL (r=.44)
Sample 2						
Separate PH, VS, and BL factors						
Unconstrained	.94	.93	.07 (.04-.09)	--	All: .69-.90	P-T (r=.53), BL-VS (r=.59), BL-PH (r=.63), PH-VS (r=.62), T-BL (r=.39), T-PH (r=.24), T-VS (r=.29), P-VS (r=.33), P-BL (r=.38), P-PH (r=.35)
Constrained	.93	.92	.07 (.05-.09)	-7.41 [5], p=.99	All: .68-.90	PH-VS = BL-VS = BL-PH (r=.61) > P-T = T-VS = P-VS = BL-VS = T-BL = BL-PH (r=.35) > T-PH (r=.21)
Separate WM and BL factors						
Unconstrained	.98	.97	.05 (.01-.08)	--	All: .69-.90	P-T (r=.51), BL-WM (r=.77), T-BL (r=.35), T-WM (r=.30), P-BL (r=.36), P-WM (r=.40)
Constrained	.98	.97	.05 (.02-.07)	4.93 [4], p=.30	All: .69-.90	BL-WM (r=.79) > P-T = T-WM = P-WM = T-BL = P-BL (r=.38)

Note. AIC = Akaike information criteria, BIC = Bayesian information criteria, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, TLI = Tucker-Lewis index. BL = Baseline painting condition. P = Parent. PH = Phonological working memory. T = Teacher. VS = Visuospatial working memory. WM = working memory.