Executive functioning rating scales: Ecologically valid or construct invalid?

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

[1]Florida State University, Department of Psychology

Abstract

Objective—Executive functions are commonly measured using rating scales and performance tests. However, replicated evidence indicates weak/non-significant cross-method associations that suggest divergent rather than convergent validity. The current study is the first to investigate the relative concurrent and predictive validities of executive function tests and ratings using (a) multiple gold-standard performance tests, (b) multiple standardized rating scales completed by multiple informants, and (c) both performance-based and ratings-based assessment of academic achievement – a key functional outcome with strong theoretical links to executive function.

Method—A well-characterized sample of 136 children oversampled for ADHD and other forms of child psychopathology associated with executive dysfunction (ages 8–13; 68% Caucasian/non-Hispanic) completed a counterbalanced series of executive function and academic tests. Parents/teachers completed executive function ratings; teachers also rated children's academic performance.

Results—The executive function tests/ratings association was modest (r=.30) and significantly lower than the academic tests/ratings association (r=.64). Relative to ratings, executive function tests showed significantly higher cross-method predictive validity and significantly better within-method prediction; executive function ratings failed to demonstrate improved within-method prediction. Both methods uniquely predicted academic tests and ratings.

Conclusion—These findings replicate prior evidence that executive function tests and ratings cannot be used interchangeably as executive function measures in research and clinical applications, while suggesting that executive function tests may have superior validity for predicting academic behavior/achievement.

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.

Executive functions are higher-order neurocognitive processes associated with regulating thoughts and behaviors by maintaining problem sets to attain future goals (Miyake et al. 2000; van der Ven et al., 2013; Wiebe et al., 2011). Deficits in executive functions are theorized to be etiologically important for a broad range of psychopathologies, including schizophrenia (Mesholam-Gately et al., 2009), depression (Joormann & Gotlib, 2008; Snyder, 2013), bipolar disorders (Bora et al., 2009), conduct disorder (Morgan & Lilienfeld, 2000) and attention-deficit/hyperactivity disorder (ADHD; Kasper et al., 2012; Willcutt et al., 2005). In addition, empirical evidence indicates that executive functions play a critical role in many important functional outcomes beginning as early as young childhood (Alloway & Alloway, 2010; Bierman et al., 2008; Gathercole et al., 2004; Moffitt et al., 2011; O'Shaughnessy et al., 2003). For example, executive function deficits negatively impact areas of social/peer and family relations (Clark et al., 2002; Gewirtz et al., 2009; Kofler et al., 2016), organizational skills (Kofler et al., 2018), occupational functioning (Barkley et al., 2008; Miller et al., 2012), and academic achievement (Best et al., 2009).

Among the diverse models of executive functions, factor analytic and theoretical work provides the most empirical support for models that include two primary executive function domains in early and middle childhood: working memory and inhibitory control (Karr et al., 2018; Lerner & Lonigan, 2014; St. Clair-Thompson & Gathercole, 2006). A third primary executive function, set shifting, is typically found in late adolescent and adult samples but may not be separable from working memory and inhibitory control in childhood (for review see Karr et al., 2018). The current study therefore focuses on working memory and inhibitory control. Working memory is a limited capacity system that involves the updating, manipulation/serial reordering, and dual-processing of internally held information for use in guiding behavior (Baddeley, 2007; Shelton et al., 2010; Unsworth et al., 2010). Neurocorrelates of working memory include the mid-lateral prefrontal cortex as well as interconnected neural networks (Nee & Jonides, 2013; Wager & Smith, 2003). Inhibitory control refers to the ability to withhold (action restraint) or suppress (action cancellation) a pre-potent behavioral response (Lewis & Carpendale, 2009). Inhibitory control is supported by diverse brain regions including the right thalamic, right superior temporal and left inferior occipital gyri, bilateral frontal, and mid-brain structures (Cortese et al., 2012).

Measurement of Executive Functions

Given the well-established associations between deficits in executive functions and adverse behavioral, academic, and occupational outcomes (Best et al., 2009; Kofler et al., 2016), accurate assessment of executive functions in childhood is critical for early detection and intervention efforts (Diamond, 2012). There are two common ways of measuring executive functions: rating scales and performance-based tests. Although their combined use has been theorized to provide a multidimensional perspective into executive functions and related behaviors (Chaytor & Schmitter-Edgecombe, 2006), recent literature indicates a strong

discord between executive function rating scales and performance-based tests (Conklin et al., 2008; Toplak et al., 2013; Vriezen & Pigott, 2002). As detailed below, concurrent and predictive validity studies consistently indicate that executive function rating scales and performance-based tests correlate weakly/non-significantly with one another (Biederman et al., 2008; Conklin et al., 2008; Nordvall et al., 2017). In other words, executive function rating scales and performance-based tests appear to show evidence for *divergent* validity rather than convergent validity. It seems psychometrically inappropriate to reify two uncorrelated tests as measures of the same underlying construct, but it is currently unclear which method is more valid for the assessment of executive functioning. That is, the weak relation between executive function rating scales and performance-based tests questions whether both can be used as reliable measurements of executive functions but cannot inform whether either method is a better indicator than the other. Additionally, as reviewed below, conclusions regarding which method(s) validly capture executive functions are further limited by methodological confounds that question the extent to which discrepant findings across studies are an artifact of systematic error variance such as shared-method bias.

Executive function performance tests

Performance-based executive function tests are computerized and/or paper-pencil tasks used to assess specific executive functions under controlled conditions (Muris et al., 2008). While support for interpreting traditional neuropsychological tests as measures of executive functioning is limited (for review, see Snyder et al., 2015), psychometric support for tests from the cognitive sciences designed specifically to measure executive functioning abilities (Snyder et al., 2015) includes replicated evidence that these tests contribute significantly and uniquely to latent estimates of global and specific executive functions (Miyake et al., 2000; St. Clair-Thompson & Gathercole, 2006; Willoughby & Blair, 2016). Support for the use of construct-valid executive function tests such as the stop-signal task and Rapport working memory tests also includes well-documented internal consistency and test-retest reliability (Alderson et al., 2008; Rapport et al., 2009; Sarver et al., 2015; Soreni et al., 2009). In addition, their concurrent and predictive validity has been supported across a range of cognitive, developmental, and clinical studies demonstrating strong convergence with ecologically valid functional outcomes (Kofler et al., 2016; Miyake et al., 2000; St. Clare-Thompson & Gathercole, 2006; Wells et al., 2018). This concurrent and predictive validity evidence includes experimental and longitudinal linkages between executive function performance tests and objective and subjective measures of inattention, hyperactivity, impulsivity, emotion recognition, reading skills, math performance, social skills, organizational skills, parent-child relationship quality, and activities of daily living (Carlson & Wang, 2007; Karalunas et al., 2017; Kofler et al., 2010, 2017; Raiker et al., 2012; Rapport et al., 2009; Wells et al., 2018).

At the same time, performance-based tests have been criticized for poor generalizability relative to rating scales because scores may reflect optimal performance under controlled conditions (Muris et al., 2008) that bear little resemblance to real world situations in which executive functions are used to guide behavior (i.e., poor external/face validity). The time cost of administering performance-based executive function tests may also be greater compared to rating scales (Toplak et al., 2013).

Executive function rating scales

Executive function rating scales are cost- and time- effective measures completed by knowledgeable observers (e.g., parents, teachers) and theorized to capture executive functions 'in the wild' – that is, executive functions as they are implemented in everyday, real-world settings (Muris et al., 2008; Toplak et al., 2013). Support for the use of executive function rating scales such as the Behavior Rating Inventory of Executive Function (BRIEF) include well-documented internal consistency and test-retest reliability (Gioia et al., 2000; Kamphaus & Reynolds, 2007). In addition, their convergent validity has been supported across multiple studies showing strong convergence with other questionnaire-based executive function assessments completed by the same informant at the same time point (Buchanan, 2016; Gross et al., 2015; McAuley et al., 2010; Muris et al., 2008). Concurrent and predictive validity evidence includes demonstrations that executive function rating scales predict theoretically-linked ratings of personality traits, internalizing and externalizing disorders, social skills, mood difficulties and academic functioning (Buchanan, 2016; Gerst et al., 2017; Gross et al., 2015; Jarratt et al., 2005; Muris et al., 2008).

At the same time, the construct validity of executive function rating scales has been questioned because they consistently correlate non-significantly (Biederman et al., 2008; Conklin et al., 2008; Mahone et al., 2002) or weakly (Gross et al. 2015; Miranda et al., 2015; Nordvall et al., 2017) with performance-based tests measuring the same construct. In addition, the content validity of executive function rating scales has been questioned, with some authors concluding that executive function rating scales measure the success of goal pursuit (Toplak et al., 2013) or externalizing behaviors rather than cognitive functioning as intended (Spiegel et al., 2017).

Executive function tests vs. rating scales

The poor correspondence between executive function rating scales and performance-based tests has frequently been interpreted to reflect the improved ecological validity of executive function rating scales (Toplak et al., 2013), particularly given the concurrent and predictive validity evidence reviewed above. In addition, the few studies that have pitted these two measurement methods head-to-head have generally found that executive function rating scales predict functional outcomes better than performance-based tests (Barkley & Murphy, 2010; Toplak, 2008), suggesting that executive function rating scales may be better measures of executive function than performance-based tests (Toplak, 2008). This conclusion warrants scrutiny, however, because to our knowledge the current evidence base relies almost exclusively on questionnaire-based outcome measures completed by the same informant at the same time point as the executive function rating scales (i.e., mono-informant, mono-method bias; Spiegel et al., 2017). In addition, the 'executive function' performance tests used in previous head-to-head studies have been uniformly criticized for poor specificity, such that they were not developed to assess executive functioning but rather gross neuropsychological functioning (for review see Snyder et al., 2015). Thus, it is unclear whether the evidence indicates that executive function rating scales are more ecologically valid than performance tests or whether these findings may be more parsimoniously attributed to mono-method, mono-informant, and/or test specificity problems. The current study addresses these limitations by using multi-informant, multi-method, and construct-

specific tests to evaluate the concurrent and predictive validity of executive function rating scales and performance-based tests.

Current Study

As reviewed above, the evidence base at this time is comprised primarily of studies exploring how well executive function rating scales or performance tests predict outcomes such as academic achievement separately, but not how they perform head-to-head when gold-standard forms of both measurement approaches are compared. To our knowledge, the lone exception to this critique is a recent study by Gerst et al. (2017), who replicated the weak-to-nonsignificant relations between executive function rating scales and performance tests ($r = .20–.25$). They also found that executive function tests showed consistent univariate relations across multiple academic tests, whereas the pattern for executive function ratings was more nuanced and measure-specific. However, despite these methodological advancements, Gerst et al. (2017) were unable to fully test for cross-domain predictive validity because they did not include rating-based academic outcome measures.

The current study builds on this work and is the first to address key limitations of the current knowledge base by investigating the relative concurrent and predictive validities of executive function rating scales and performance-based tests via the inclusion of (a) multiple counterbalanced tests developed specifically to assess executive functioning abilities as defined in the cognitive sciences (e.g., Baddeley, 2007; Miyake & Friedman, 2012); (b) multiple standardized, gold standard executive function rating scales completed by multiple informants (parents and teachers); (c) both performance-based and questionnaire-based assessment of academic achievement – a key functional outcome domain with strong theoretical links to executive function as reviewed below; and (d) a large and diverse sample of children oversampled for ADHD and other forms of child psychopathology associated with executive dysfunction (e.g., Kasper et al., 2012). Consistent with prior work, we predicted that executive function rating scales and executive function performance tests would be weakly or non-significantly associated with one another (e.g., Toplak et al., 2013). We also predicted evidence consistent with mono-method bias. That is, we expected performance-based executive function tests to predict performance-based academic tests and questionnaire-based executive function measures to predict questionnaire-based academic measures. Strong support for the predictive validity of executive function tests and/or ratings would require cross-method prediction (i.e. performance-based executive function tests predicting questionnaire-based academic measures and/or questionnaire-based executive function measures predicting performance-based academic tests). No predictions were made regarding cross-method predictive validity due to limitations of the available evidence as described above.

Methods

Participants

The sample comprised 136 children (48 girls, 88 boys), aged 8 to 13 years (M = 10.34, SD = 1.53) from the Southeastern United States recruited through community resources from 2015–2018 for participation in a larger study of the neurocognitive mechanisms underlying

pediatric attention and behavioral problems. Institutional Review Board approval was obtained/maintained, and all parents and children gave informed consent/assent. Sample ethnicity was mixed with 92 Caucasian Non-Hispanic (67.6%), 15 Black (11.0%), 17 Hispanic (12.5%), 9 multiracial (6.6%), and 3 Asian children (2.2%; Table 1).

All children and caregivers completed a comprehensive evaluation that included detailed, semi-structured clinical interviewing and multiple norm-referenced parent and teacher questionnaires. A detailed account of the comprehensive psychoeducational evaluation can be found in the larger study's preregistration: https://osf.io/nvfer/. The final sample was composed of 136 children, including an oversampling of children with ADHD: 42 children with ADHD, 44 children with ADHD and common comorbidities (52.2% anxiety, 9.1% depressive, 20.5% oppositional defiant, 18.2% autism spectrum disorders), 27 with common clinical diagnoses but not ADHD (63.0% anxiety, 14.8% depressive, 7.4% oppositional defiant, 25.9% autism spectrum disorders),[1] and 23 neurotypical children. Psychostimulants ($N_{prescribed}$=29) were withheld 24-hours for neurocognitive testing. Psychoeducational evaluations were provided to parents. Children were excluded for gross neurological, sensory, or motor impairment; non-stimulant medications that could not be withheld for testing; or history of seizure disorder, psychosis, or intellectual disability.

Procedures

Executive function testing occurred as part of a larger battery that involved 1–2 sessions of approximately 3 hours each. 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. 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).

Performance-Based Executive Function Tests

Rapport working memory tests—The Rapport et al. (2009) computerized working memory tests and their administration instructions are identical to those described in Kofler et al. (2018). 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). Six trials per set size were administered in randomized/unpredictable order (3–6 stimuli/trial; 1 stimuli/second) as recommended (Kofler et al., 2016). Five practice trials were administered before each task (80% correct required). Task duration was approximately 5 (visuospatial) to 7 (phonological) minutes. Partial-credit unit scoring (i.e., stimuli correct per trial) was used to index overall working memory performance at each set size 3–6 was used as recommended (Conway et al., 2005).

[1]Exploratory analyses indicated that excluding children with ASD did not alter the pattern, significance levels, and interpretation of results.

In the Rapport phonological working memory test, children were presented a series of jumbled numbers and a letter (1 stimuli/second). 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).

In the Rapport visuospatial working memory test, children were shown nine squares arranged in three offset vertical columns. A series of 2.5 cm dots were presented sequentially (1 stimuli/second); 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.

Stop-signal inhibitory control—The stop-signal test and administration instructions are identical to those described in Alderson et al. (2008). Psychometric evidence includes high internal consistency (α = .80; Kofler et al., 2019) and three-week test–retest reliability (.72), as well as convergent validity with other inhibitory control measures (Soreni et al., 2009). Go-stimuli are displayed for 1000 ms as uppercase letters X and O positioned in the center of a computer screen (500 ms interstimulus interval; total trial duration = 1500 ms). Xs and Os appeared with equal frequency throughout the experimental blocks. A 1000 Hz auditory tone (i.e., stop-stimulus) was presented randomly on 25% of trials. Stop-signal delay (SSD) —the latency between presentation of go- and stop-stimuli—is initially set at 250 ms and dynamically adjusted ± 50 ms contingent on participant performance. Successfully inhibited stop-trials are followed by a 50 ms increase in SSD, and unsuccessfully inhibited stop-trials are followed by a 50 ms decrease in SSD. All participants completed two practice blocks and four consecutive experimental blocks of 32 trials per block (24 go-trials, 8 stop-trials per block). SSD was selected based on conclusions from recent meta-analytic reviews that it is the most direct measure of inhibitory control in stop-signal tasks that utilize dynamic SSDs, given that SSDs change systematically according to inhibitory success or failure (Alderson et al., 2007; Lijffijt et al., 2005). Higher SSD scores indicate better inhibitory control.

Go/no-go inhibitory control—The go/no-go is a response inhibition task in which a motor response must be executed or inhibited based on a stimulus cue (Bezdjian et al. 2009). Psychometric evidence includes high internal consistency (α = .95) as well as convergent validity with other inhibitory control measures (Kofler et al., 2019). Children were presented a randomized series of vertical (go stimuli) and horizontal (no-go stimuli) rectangles in the center of a computer monitor (2000 ms presentation, jittered 800–2000 ms ISI to minimize anticipatory responding). They were instructed to quickly click a mouse button each time a vertical rectangle appeared, but to avoid clicking the button when a horizontal rectangle appeared. A ratio of 80:20 go:no-go stimuli was selected to maximize prepotency (Kane & Engle 2003; Unsworth & Engle 2007). Children completed a 10 trial practice (80% correct required) followed by 4 continuous blocks of 25 trials each. Commission errors reflect failed inhibitions (i.e., incorrectly responding to no-go trials), and served as the primary index of inhibitory control during each of the four task block. Lower scores reflect better inhibition.

Dependent variables: Executive function performance—Task impurity was controlled by computing a Bartlett maximum likelihood estimate based on the intercorrelations among task performance scores (i.e., component score; DiStefano et al., 2009), which combined the 8 working memory (set sizes 3–6) and 8 inhibitory control (blocks 1–4) performance variables into a single executive function performance composite score (28% variance accounted; Supplementary Table 1). This PCA-derived component score provides an estimate of reliable, component-level variance attributable to domain-general executive functioning. This formative method for estimating executive functioning was selected because (a) such methods have been shown to provide higher construct stability relative to confirmatory/reflective approaches (Willoughby et al., 2016); and (b) estimating executive functioning at the construct-level rather than measure-level was expected to maximize associations between component estimates of executive function tests and executive function ratings to the extent that these methods are tapping the same underlying construct as hypothesized. Conceptually, this process isolates reliable variance across estimates of executive function by removing task-specific demands associated with nonexecutive processes, time-on-task effects via inclusion of four blocks per task, and non-construct variance attributable to other measured executive and non-executive processes (e.g., short-term memory load). This executive function performance component score was used in all analyses below. Higher scores reflect better executive functioning.

Questionnaire-Based Executive Function Measures

Behavior Rating Inventory of Executive Function (BRIEF)—The BRIEF (Gioia et al., 1996) parent and teacher forms are each 86-item scales that assess executive function impairments in children ages 5–18. Raw scores are converted to age- and sex-specific T-scores based on the national standardization sample (N=1,419 per form). Parent and teacher Global Executive Composite T-scores served as the primary indices of executive function functioning at home and school, respectively. Psychometric support for these subscales include high internal consistency (α = .89–.98), test-retest reliability (r = .76–.91), and expected correspondence with other informant-based ratings of executive functions (Gioia et al., 1996; Sullivan & Riccio, 2006).

Behavior Assessment System for Children (BASC-2/3)—The BASC-2/3 (Reynolds & Kamphaus, 2004, 2015) parent and teacher forms are 139–175 item scales that assess internalizing and externalizing behavior problems in children ages 2–21. Raw scores are converted to age- and sex-specific T-scores based on the national standardization sample (N = 1,800 per form). Parent and teacher Executive Function subscale T-scores served as the primary indices of executive function functioning at home and school, respectively. Psychometric support for the Executive Functioning subscale includes high internal consistency (α=.85–.95), test-retest reliability (r = .80–.92), and expected correspondence with other informant-based ratings of executive functions (Reynolds & Kamphaus, 2004, 2015; Sullivan & Riccio, 2006).

Dependent variables: Executive function ratings—As described above, a Bartlett maximum likelihood component score was computed by combining the 2 parent and 2 teacher (BRIEF and BASC for each informant) executive functioning ratings (48% variance

explained; loadings = .43–.89). This executive function ratings component score was used in all analyses below. Scores were multiplied by −1 so that higher scores reflect better executive functioning for both ratings and performance tests.

Predictive Validity Outcome: Academic Functioning

Academic functioning was selected as the predictive validity outcome given its centrality in children's lives, association with social and occupational functioning both cross-sectionally and longitudinally (Kuriyan et al., 2013; Walker, & Nabuzoka, 2007; Wentzel, 1991), and strong theoretical and empirical links with executive functioning (e.g., Ahmed et al., 2018; Samuels et al., 2016). Several studies have demonstrated the contributions of executive functions to a wide range of academic achievement outcomes (Bull et al., 2008; Hitch et al., 2001; Miller & Hinshaw, 2010). For example, executive functions have been associated with academic skills involving written language (Alloway et al., 2005), oral language (McInnes et al., 2003), reading and math attainment (Raghubar et al., 2010; Sesma et al., 2009; Swanson & Kim, 2007; Thorell, 2007; Wåhlstedt et al., 2009), reading comprehension (Kieffer et al., 2013; Kofler, Wells et al., 2018), science (Gathercole et al., 2004; St. Clair-Thompson & Gathercole, 2006), and following instructions (Jaroslawska et al., 2016).

Performance-based academic tests—The Kaufman Test of Educational Achievement (KTEA-2/3; Kaufman & Kaufman, 2004, 2014) was used to assess academic achievement (α = .92–.99; 1–2 week test-retest = .80–.96). The Comprehensive Academic Composite/Academic Skills Battery Composite Score provides an overall index of academic functioning. Standard scores were obtained by comparing performance to the nationally representative standardization sample (N = 3,000) according to age. Higher scores indicate greater academic achievement.

Questionnaire-based academic measures—The Academic Performance Rating Scale (APRS; DuPaul et al., 1991) is a 19-item measure completed by each child's teacher to assess academic functioning (2-week test–retest = .93 to .95, α = .94 to .95). T-scores for overall academic functioning (APRS Total Score) were computed based on the standardization sample (N = 487) according to age and sex. Higher scores indicate greater academic success and productivity.

Dependent variable: academic functioning—The norm-referenced KTEA and APRS scores were converted to z-scores based on the current sample to match the scaling of the EF ratings and performance test component scores. Higher scores reflect greater academic functioning.

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

All children were administered the Verbal Comprehension Index of a Wechsler (2011, 2014) scale (WISC-IV, WISC-V). Hollingshead (1975) SES was estimated based on caregiver(s)' education and occupation.

Data Analysis Overview

The current study tested for cross-method concurrent and predictive validity of performance-based and questionnaire-based methods for measuring executive function, both relative to each other and relative to academic achievement – a key ecologically-valid outcome with strong theoretical links to executive function. In the first analytic Tier, we examined construct-level associations between executive function ratings and performance tests, and used the test of dependent correlations as implemented in the R package cocor (Diedenhofen & Musch, 2015) to test whether the component EF ratings/tests correlation was similar to the academic functioning ratings/performance correlation. That is, to what extent is the association between executive function tests and ratings comparable to the association seen for other important aspects of children's functioning when assessed across methods?

In Tier 2, we tested the extent to which the construct-level estimates (i.e., component scores) of executive function ratings and performance tests (a) uniquely predicted mono-method academic ratings and test performance (i.e., ratings predicting ratings, tests predicting tests); and (b) provided strong predictive validity evidence by uniquely predicting cross-method academic functioning (i.e., ratings predicting tests, tests predicting ratings). This involved creating a path model in which the EF ratings and EF performance test component scores (allowed to correlate) were each modeled to predict academic ratings and academic performance (which were also allowed to correlate). We then tested whether these relations differed significantly in magnitude (e.g., does one method for assessing executive functions show stronger predictive validity evidence than the other?). This involved constraining pathways to be equal and assessing whether doing so significantly reduced model fit; a significant Wald χ^2 test indicates that the pathways are significantly different in magnitude (i.e., that constraining them to be equal significantly degrades model fit).

Age, sex, medication status (no/yes), and ADHD status (no/yes) were covaried in all analyses; the pattern, statistical significance, and interpretation of results are unchanged in exploratory analyses in which these covariates were removed.

Results

Power Analysis

A series of Monte Carlo simulations were run using Mplus7 (Muthén & Muthén, 2002, 2012) to estimate the power of our proposed path model for detecting significant relations between executive functions and academic achievement, given power (1-β) .80, α=.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 utilizing executive function rating scales and performance tasks (Alderson et al., 2010; Jarratt et al., 2005; Sullivon & Riccio, 2006) were imputed to delineate the proposed path model. Based on these parameters, the smallest sample size necessary for this study was N=53 to detect relations between executive functions and academic achievement of β=.30 or greater. Thus, our study (N=136) is sufficiently powered to address the study's primary aims.

Preliminary Data Analysis and Descriptive Statistics—Primary analyses were conducted using M-Plus 7.4 (Muthén & Muthén, 2012). Data-screening procedures ensured that assumptions regarding multivariate normality, absence of outliers, linearity, and homogeneity of variance-covariance matrices were not violated (Tabachnick & Fidell, 2013). Missing data rates were low (1.5%) and were missing completely at random (Little's MCAR test: χ^2 [211] = 70.9, p >.99). These data were therefore imputed using the expectation-maximization (EM) algorithm (Shafer, 1997). Task data from subsets of the current battery have been reported for subsets of the current sample to examine conceptually unrelated hypotheses in Kofler et al. (2018, 2019). Data from the executive function rating scales, academic performance tests, and academic rating scales have not been previously reported. Descriptive statistics for all subtests are shown in Table 1. All univariate skewness values and kurtosis values were less than an absolute value of 1.0 for primary and secondary outcomes and predictors.

Tier 1: Concurrent Validity of Executive Function Tests and Ratings—As shown in Table 2, executive function performance and ratings showed a significant, albeit modest, zero-order association (r=.30, p<.001), whereas academic performance and ratings showed the large magnitude association expected for two methods that assess the same underlying construct (r=.63, p<.001). The test of dependent correlations indicated that the difference between these associations was significant (r = .33, z=4.01, p<.0001, 95%CI = .17–.50), indicating that executive function tests and ratings do not show the expected level of correspondence relative to expectations based on associations between tests and ratings of other important aspects of childhood functioning. To wit, squaring the bivariate correlations indicated that executive function performance/ratings share only 9% of their variance, relative to 40% variance shared between academic performance and ratings. These findings are particularly striking given our use of construct-level estimates of performance tests and ratings that were each based on multiple criterion measures, but are consistent with prior work (reviewed above) and provide additional evidence that executive function tests and ratings appear to be weakly correlated despite both being reified as measures of 'executive function.'[2] However, as noted above, this divergent validity evidence does not inform whether one (or either) method is more valid for assessing executive functions. Thus, the Tier 2 analyses assessed their relative predictive validity for understanding a critical developmental task of childhood (academic functioning) known to depend on executive functioning abilities.

Tier 2: Predictive Validity of Executive Function Tests and Ratings—As shown in Figure 1, the mono-informant hypothesis was supported, such that the EF ratings component predicted academic ratings (β=.37, p<.001) and the EF performance test component predicted academic performance tests (β=.64, p<.001). There was also support

---

[2]In addition to testing the latent/component-level associations between executive function tests and ratings, we also explored this via an exploratory factor analysis that included the 4 tests and 4 rating scales. Results indicated that a 2-factor solution was preferred based on both parallel analysis and eigenvalue > 1 (orthogonal rotation; 57.16% variance accounted for). The 2 factors broke down into a Test factor (all 4 EF tests loaded .66–.74, all 4 EF ratings loaded minimally at .02–.23) and a Ratings factor (all 4 EF ratings loaded .69–.84, all 4 EF tests loaded minimally at .01 to .29), providing additional evidence to suggest that these methods may be assessing different constructs. Of note, this factor analysis was not planned *a priori* but was added during the peer review process; results should therefore be considered exploratory.

for the predictive validity of both measurement methods, such that EF ratings uniquely predicted academic performance tests after accounting for EF performance (β=.20, p=.005) and EF performance tests uniquely predicted academic ratings after accounting for EF ratings (β=.47, p<.001). Collectively, the EF performance tests and ratings explained a large proportion of the variance in children's academic achievement as assessed by both academic performance tests (R^2=.42, p<.001) and ratings (R^2=.40, p<.001).

Results of the Wald χ^2 tests indicated that EF performance tests showed better cross-method predictive validity (β=.47 vs. .20; p=.04) and also showed better mono-method prediction (β=.64 vs. .37; p=.01) relative to EF ratings. In addition, EF performance tests were stronger predictors of academic performance (β=.64 vs. .20; p<.001), whereas EF tests and ratings predicted academic ratings similarly (β=.47 vs. .37; p=.38). EF performance tests showed uniformly strong prediction of both academic tests and ratings (β=.47–.64; p=.18), whereas cross-method prediction was significantly lower than mono-method prediction for EF ratings (β=.37 vs. .20; p=.001).

Taken together, the Tiers 1–2 findings indicate that EF performance tests and ratings show modest, albeit significant, construct-level associations with each other that fall well below expectations based on the magnitude of the test/ratings association for other important aspects of childhood functioning (i.e., only 9% variance shared between component estimates of two methods intended to assess the same construct). Both methods provide unique prediction for understanding children's academic functioning, with EF performance tests showing stronger cross-method predictive validity and outperforming EF ratings for predicting academic performance tests. In contrast, EF ratings failed to outperform EF performance tests for predicting academic ratings and, consistent with the mono-informant bias hypothesis, showed a small magnitude association with academic performance tests that was significantly lower than its association with academic ratings.

Discussion

The current study was the first to employ a multi-method, multi-informant/multi-test approach with component estimation to assess for evidence of construct-level concurrent and predictive validity between executive function rating scales and performance-based tests. Additional strengths include the relatively large, clinically evaluated sample and multi-method assessment of academic functioning – a key predictive validity outcome with strong empirical and theoretical links with executive function. Overall, we replicated prior findings of modest to weak associations between executive function tests and ratings (Gross et al., 2015; Miranda et al., 2015; Nordvall et al., 2017) and extended those findings by showing that, even at the construct level, these associations fall well below expectations for cross-method associations between measures intended to assess the same construct (i.e., only 9% variance shared between two executive function assessment methods vs. 40% shared between academic tests/ratings). Despite this apparent divergent validity evidence, both methods were important for predicting children's academic functioning (Gerst et al., 2017), albeit to different degrees and with a more nuanced pattern of findings that overall provided greater support for the construct and predictive validity of executive function performance tests as discussed below.

With regard to concurrent validity, we found a significant but modest relation between construct-level estimates of executive function performance tests and rating scales (r=.30). This finding was highly consistent with prior literature showing modest to non-significant measure-level relations between these two putative methods for assessing the same construct in child, adolescent, and adult populations (Biederman et al., 2008; Conklin et al., 2008; Mahone et al., 2002), and extended previous findings by demonstrating that this relation remained small when controlling for mono-measure and task/measure-level error. A parsimonious explanation for this modest association could be that executive function tests and ratings are assessed using different methods completed in different settings across different time periods (Podsakoff et al., 2003). That is, executive function performance tests are typically completed by children in laboratory or clinic settings over a relatively short duration (i.e., typically 5–10 minutes per test), whereas executive function ratings are completed by teachers/parents based on children's behavior in the classroom/at home over a longer duration (i.e., most rating forms ask informants to consider the last several weeks or months; Gioia et al., 1996; Reynolds & Kamphaus, 2004, 2015; DuPaul et al., 1991). This explanation appears unlikely, however, given the robust association between academic performance tests and ratings (r=.63) despite highly similar discrepancies in terms of setting, informant/examinee, and timeframe.

An alternative explanation that has been proposed recently is that, despite both being reified as measures of 'executive function,' executive function performance tests and ratings are measuring fundamentally different underlying constructs (Spiegel et al., 2017; Toplak et al., 2013). Whereas modern executive function tests were developed based on cognitive models of executive function (e.g., Miyake & Friedman, 2012), such that construct-level estimates based on multiple tests appear to reliably and validly capture executive function-specific cognitive abilities (e.g., Schmiedek et al., 2015; Toplak et al., 2013), executive function ratings have been hypothesized to instead reflect mental constructs involving the success of goal pursuit (Toplak et al., 2013) and/or non-mental constructs such as externalizing behavior problems (Spiegel et al., 2017). The current study contributed to this debate by replicating the overall weak association between executive function performance tests and ratings at the construct level and providing the first fully-crossed test of predictive validity relative to a key functional outcome domain with strong theoretical links to executive function (i.e., academic achievement, Ahmed et al., 2018; Samuels et al., 2016).

It seems psychometrically inappropriate to reify two weakly correlated assessment tools as measures of the same underlying construct, but it has been unclear which method is more valid for the assessment of executive functioning (Toplak et al., 2013). Along those lines, of primary interest in the current study was the extent to which performance-based and questionnaire-based measures of executive function predict an important childhood functional outcome known to depend on executive function abilities (academic achievement). Overall, results indicated that both measurement methods demonstrated unique prediction for children's academic functioning. However, executive function performance tests provided stronger cross-method predictive validity and outperformed executive function ratings for predicting academic performance tests. In addition, executive function ratings failed to outperform executive function performance tests for predicting academic ratings. This pattern of results is generally consistent with recent findings by Gerst

et al. (2017) who reported that executive function performance tests showed generally higher univariate associations with math achievement (r=.49–.57 for executive function performance tests versus .29–.42 for ratings), although similar associations were found for reading (r=.32–.55 versus .38–.55).

In contrast, the current study's findings of superior predictive validity for executive function performance tests vs. ratings appears inconsistent with several studies finding that executive function ratings predicted functional outcomes better than executive function performance tests (Barkley & Murphy, 2010; Gioia et al., 2000; Gross et al., 2015; Kamradt et al., 2014; Mahone & Hoffman, 2007; Toplak, 2008). A critical distinction between the current study and studies failing to support the predictive validity of executive function performance tests is that most head-to-head studies used questionnaire-based outcome measures completed by the same informant at the same time point as the executive function questionnaires (i.e., mono-method, mono-informant bias) and performance tests that were developed to assess gross neuropsychological dysfunction rather than executive functioning specifically (for review see Snyder et al., 2015). Further research is needed to assess the predictive validity of executive function tests and ratings relative to other important outcomes known to depend at least in part on executive function abilities, such as phonological awareness (Allan et al., 2015), social relations (Clark et al., 2002), organizational skills (Kofler et al., 2018), and occupational functioning (Barkley, Murphy, & Fischer, 2008).

Limitations

The current study was the first to comprehensively assess the concurrent and predictive validities of executive function rating scales and performance-based tests using a fully-crossed design (i.e., tests- and ratings-based assessment of all predictors and outcomes), construct-level estimates of both assessment methods derived from multi-method, multi-informant/multi-test indicators, and a relatively large sample of carefully phenotyped children. Despite these methodological refinements, the following limitations must be considered when interpreting results. This study was cross-sectional and the sample was comprised primarily of clinically-referred children ages 8 to 13 years, including an oversampling of children with ADHD and other forms of child psychopathology. Oversampling for ADHD and other forms of child psychopathology was considered a strength given that (a) these disorders have well-documented associations with executive dysfunction (e.g., Kasper et al., 2012), and (b) generalizability is improved because the sample is more representative of the types of children for whom these tests and ratings are administered in clinical practice (i.e., children referred for attention, learning, behavior, or emotional concerns). Given that patterns of deficits in executive functioning are highly variable among children with ADHD (Castellanos & Tannock, 2002; Kofler et al., 2019), the current study's oversampling for children with ADHD was expected to maximize executive function tests/ratings associations by providing a broader range of scores, particularly when combined with inclusion of (a) other conditions associated to greater or lesser extents with difficulties in the assessed domains, and (b) neurotypical children more likely to exhibit strengths in the assessed domains. Nonetheless, the inclusion of children without ADHD may reduce specificity of the findings to ADHD, just as the oversampling of children with ADHD may reduce generalizability to the broader population of children. Additional cross-

sectional and longitudinal studies of typically developing, community, and clinical samples across various developmental periods are necessary to assess the generalizability of these findings.

Next, our academic ratings outcome was limited to a single behavioral measure (i.e., teacher APRS ratings), whereas academic test scores were based on a composite of multiple K-TEA subtests. Future work should include academic ratings from multiple informants to more precisely match the academic test outcome, despite the robust association between the two indicators found in this study (r=.63). Similarly, the current study was limited to a single outcome domain (academic functioning); studies assessing performance-based and ratings-based outcomes across specific academic domains (e.g., reading vs. mathematics) or important predictive validity domains known to depend on executive functions (e.g., peer and family functioning, organizational skills) will be critical for determining the generalizability of the current findings. Similarly, future work may care to replicate and extend the current findings using specific executive functions (e.g., working memory, inhibitory control), using multiple measures of performance-based tests and rating scales for each examined construct.[3] Additionally, ecological validity concerns remain despite a growing body of evidence that modern executive function tests predict a host of 'real world' outcomes (e.g., Kofler et al., 2018; Wells et al., 2018); technology-driven methods such as virtual reality-based executive function assessment may help to balance the internal validity gained through careful, laboratory-based control with the external/ecological validity gained through field investigations (Jaroslawska et al. 2016; Jovanovski et al., 2012; Parsons et al., 2017; Renison et al., 2012).

Similarly, at first glance the modest association between executive function ratings and tests could be explained by the former's coverage of a wide range vs. the latter's coverage of a more narrow range of executive functions. That is, the current study's performance tests focused on working memory and inhibition, whereas the executive function ratings focused on a broader range of executive functions and their behavioral outcomes (e.g., shifting, organization, planning, emotional control). However, this conclusion seems unlikely given our use of component scores to capture domain-general executive functioning abilities, combined with replicated evidence that these tests are capturing reliable variance associated with the wide range of behaviors that the EF ratings are intended to capture, including evidence that these tests capture variance related to working memory and inhibition, as well as organization and planning (Kofler et al., 2017), set shifting (Irwin et al., 2019), and emotional control (Groves et al., 2020). Thus, the evidence indicates strongly that our executive function test and ratings composite are likely covering similarly wide ranges of abilities.

---

[3]We considered repeating the study's analyses separately for tests vs. ratings of working memory, and tests vs. ratings of inhibition. However, (1) the BASC does not include specific EF subscales, which precludes our ability to use multiple measures for each construct, and (2) inspection of the items that load onto the BRIEF Working Memory and Inhibit subscales raise significant concerns about interpreting them as measures of these specific executive functions rather than as externalizing behavior scales as argued previously (Spiegel et al., 2017). That is, these subscales' item sets appear almost complete redundant with the DSM-5 (APA, 2013) ADHD Attention Problems and Hyperactivity/Impulsivity diagnostic criteria, respectively (e.g., the BRIEF Working Memory subscale contains 10 items, 7 of which appear to have just minor wording differences from DSM-5 Attention Problems symptoms).

In addition, it is possible that the academic tests/ratings association (r=.63) was an unrealistically high bar for comparison given that this relation may be more transparent and direct. In other words, classroom performance and academic skills are repeatedly and quantitatively assessed by teachers, often using criteria similar to the skills measured on academic achievement tests. In contrast, observers do not have repeated, direct quantitative measurements of neurocognitive functioning (or most other realms of human behavior), making the association less transparent and potentially making it more difficult to find convergence between measurement methods. Future work with cross-method assessment of other areas of functioning is clearly needed to determine a reasonable bar for concluding that different methods are assessing the same vs. different constructs.

Finally, at first glance our conclusion that r=.30 provides evidence that executive function tests and ratings cannot be used interchangeably may seem at odds with the common finding that different executive function tests often correlate with each other in this range despite contributing uniquely to the latent measurement of executive functions. Correlations of this magnitude among individual executive function tests highlight the task impurity problem, such that the majority of variance in a single executive function test is not attributable specifically to the executive function the test is designed to measure (Snyder et al., 2015). In this vein, it is tempting to conclude that executive function tests and ratings may both be measuring executive functioning, albeit with similar levels of impurity that prevent single tests or ratings from being interpreted as indicators of executive functioning in applied settings. However, the exploratory factor analysis was consistent with our overall conclusions, and indicated that the executive function tests and ratings loaded cleanly onto distinct factors. In addition, the modest associations in the current study were based on construct-level, maximum likelihood components that explicitly accounted for task/measurement impurity by modeling reliable variance shared across multiple executive function tests and across multiple executive function ratings. That is, whereas low individual measure-level associations may reflect measurement unreliability, construct-level associations among executive functions tend to be much higher (e.g., r=.63 between working memory and inhibition in the seminal Miyake et al., 2000 paper).

Clinical and Research Implications

Returning to the study's primary question: Are executive function tests and rating scales measures of the same underlying construct? Our conclusion is "probably not." On one hand, the tests and ratings components correlated only modestly, and the tests and ratings loaded separately and cleanly in factor analysis. On the other hand, (1) the tests and ratings each independently predicted both academic outcomes – albeit with stronger predictive validity evidence for the executive function tests; and (2) the tests/ratings correlation was significant – albeit at a magnitude where others have concluded that instruments cannot be used interchangeably as measures of the same construct (Redick & Lindsey, 2013). That is, while the correlation between executive function tests and ratings was significantly greater than zero, it was still weaker than would be expected for measures of the same underlying executive function construct (Redick & Lindsey, 2013) and interestingly enough was identical to the r=.30 that Meehl (1990) characterized as reflective of a "crud factor" in psychological research (p. 125). For comparison, it is striking that the executive function

components correlated significantly higher with the academic outcomes than they did with each other.[4] The executive function and academic measures are not assumed to reflect the same underlying construct, yet their correlations are higher than the correlation between executive function tests and ratings. This, then, is the puzzle: Why would two measures of the same construct correlate less strongly with each other than with measures of putatively different constructs? (Redick & Lindsey, 2013). It is possible that they reflect cognitive vs. behavioral manifestations of executive functions (or abilities vs. skills, albeit with neurocognitive executive function abilities accounting for < 10% of the variance in executive function behaviors), or a conceptual grouping (i.e., a conceptual category rather than a construct). Our current view is that they are likely distinct yet correlated constructs, with neurocognitive executive function abilities playing a small but significant role in the behaviors indexed by these questionnaires, and confusion resulting in part from different research traditions adopting the same term for constructs that are not related to the degree that one would expect based on their shared reification. We speculate that adopting alternative terms for the behavioral ratings such as 'organization, time management, and planning skills' (OTMP skills; Abikoff et al., 2013) to refer to the specific behaviors/skills of interest will help to resolve this conflict while also promoting additional research by highlighting that there is a lot more than executive functioning abilities that are important for understanding 'executive function' behaviors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Ahmed SF, Tang S, Waters NE, & Davis-Kean P (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. Journal of Educational Psychology, 111(3), 446.

Alderson RM, Rapport MD, & Kofler MJ (2007). ADHD and behavioral inhibition: A meta-analytic review of the stop-signal paradigm. Journal of Abnormal Child Psychology, 35, 745–758.

Alderson RM, Rapport MD, Sarver DE, & Kofler MJ (2008). ADHD and behavioral inhibition: A re-examination of the stop-signal task. Journal of Abnormal Child Psychology, 36, 989–998.

Alderson RM, Rapport MD, Hudec KL, Sarver DE, & Kofler MJ (2010). Competing core processes in attention-deficit/hyperactivity disorder (ADHD): Do working memory deficiencies underlie behavioral inhibition deficits?. Journal of Abnormal Child Psychology, 38(4), 497–507.

Allan DM, Allan NP, Lerner MD, Farrington AL, & Lonigan CJ (2015). Identifying unique components of preschool children's self-regulatory skills using executive function tasks and continuous performance tests. Early Childhood Research Quarterly, 32, 40–50.

---

[4]Executive function/academic correlations are shown in Table 2. Tests of dependent correlations indicate that the EF-academic correlation is significantly stronger than the EF test-ratings correlations in 3 of the 4 comparisons (r=.46–.55 vs. .30; all p<.04); the lone exception is that the EF ratings-academic tests correlation is not significantly larger than the executive function tests-ratings correlation (r=.36 vs. .30; p=.43).

Allan NP, & Lonigan CJ (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. Developmental psychology, 47(4), 905.

Alloway TP, & Alloway RG (2010). Investigating the predictive roles of working memory and IQ in academic attainment. Journal of Experimental Child Psychology, 106(1), 20–29.

Alloway TP, Gathercole SE, Adams AM, Willis C, Eaglen R, & Lamont E (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. British Journal of Developmental Psychology, 23(3), 417–426.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Baddeley A (2007). Working memory, thought, and action. Oxford, UK: Oxford University Press.

Barkley RA, Murphy KR, & Fischer M (2008). ADHD in adults: What the science tells us. New York: Guilford.

Barkley RA, & Murphy KR (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. Archives of Clinical Neuropsychology, 25(3), 157–173.

Best JR, Miller PH, & Jones LL (2009). Executive functions after age 5: Changes and correlates. Developmental Review, 29(3), 180–200.

Bezdjian S, Baker LA, Lozano DI, & Raine A (2009). Assessing inattention and impulsivity in children during the Go/NoGo task. British Journal of Developmental Psychology, 27(2), 365–383.

Bierman KL, Nix RL, Greenberg MT, Blair C, & Domitrovich CE (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. Development and Psychopathology, 20(3), 821–843.

Biederman J, Petty CR, Fried R, Black S, Faneuil A, Doyle AE, … & Faraone SV (2008). Discordance between psychometric testing and questionnaire-based definitions of executive function deficits in individuals with ADHD. Journal of Attention Disorders, 12(1), 92–102.

Bora E, Yucel M, & Pantelis C (2009). Cognitive endophenotypes of bipolar disorder: A meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. Journal of Affective Disorders, 113(1–2), 1–20.

Buchanan T (2016). Self-report measures of executive function problems correlate with personality, not performance-based executive function measures, in nonclinical samples. Psychological Assessment, 28(4), 372.

Bull R, Espy KA, & Wiebe SA (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental Neuropsychology, 33(3), 205–228.

Carlson SM, & Wang TS (2007). Inhibitory control and emotion regulation in preschool children. Cognitive Development, 22(4), 489–510.

Chaytor N, Schmitter-Edgecombe M, & Burr R (2006). Improving the ecological validity of executive functioning assessment. Archives of Clinical Neuropsychology, 21(3), 217–227.

Clark C, Prior M, & Kinsella G (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. Journal of Child Psychology and Psychiatry, 43, 785–796.

Conklin HM, Salorio CF, & Slomine BS (2008). Working memory performance following paediatric traumatic brain injury. Brain Injury, 22(11), 847–857.

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, & Engle RW (2005). Working memory span tasks: Methodological review and user's guide. Psychonomic Bulletin & Review, 12, 769–86.

Cortese S, Kelly C, Chabernaud C, Proal E, Di Martino A, Milham MP, & Castellanos FX (2012). Toward systems neuroscience of ADHD: A meta-analysis of 55 fMRI studies. American Journal of Psychiatry, 169(10), 1038–1055.

Diamond A (2012). Activities and programs that improve children's executive functions. Current Directions in Psychological Science, 21(5), 335–341.

Diedenhofen B & Musch J (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. PLoS ONE, 10(4): e0121945.

DiStefano C, Zhu M, & Mindrila D (2009). Understanding and using factor scores: Considerations for the applied researcher. Practical Assessment, Research & Evaluation, 14(20), 1–11.

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

DuPaul GJ, Rapport MD, & Perriello LM (1991). Teacher ratings of academic skills: The development of the academic performance rating scale. School Psychology Review, 20, 284–300.

Gathercole SE, Pickering SJ, Knight C, & Stegmann Z (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 18(1), 1–16.

Gerst EH, Cirino PT, Fletcher JM, & Yoshida H (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. Child Neuropsychology, 23(4), 381–407.

Gewirtz S, Stanton-Chapman TL, & Reeve RE (2009). Can inhibition at preschool age predict attention-deficit/hyperactivity disorder symptoms and social difficulties in third grade?. Early Child Development and Care, 179(3), 353–368.

Gioia GA, Andrwes K, & Isquith PK (1996). Behavior rating inventory of executive function-preschool version (BRIEF-P). Odessa, FL: Psychological Assessment Resources.

Gioia GA, Isquith PK, Guy SC, & Kenworthy L (2000). BRIEF: Behavior rating inventory of executive function. Psychological Assessment Resources.

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

Gross AC, Deling LA, Wozniak JR, & Boys CJ (2015). Objective measures of executive functioning are highly discrepant with parent-report in fetal alcohol spectrum disorders. Child Neuropsychology, 21(4), 531–538.

Groves Nicole B., Kofler Michael J., Wells Erica L., Day Taylor N., and Chan Elizabeth SM. "An Examination of Relations Among Working Memory, ADHD Symptoms, and Emotion Regulation." Journal of Abnormal Child Psychology (2020): 1–13.

Hitch GJ, Towse JN, & Hutton U (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. Journal of Experimental Psychology: General, 130(2), 184.

Hollingshead AB (1975). Four factor index of social status.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Ibm, C. R. (2017). IBM SPSS Statistics for Windows, Version Q3 25.0. Armonk, NY: IBM Corp.

Jaroslawska AJ, Gathercole SE, Logie MR, & Holmes J (2016). Following instructions in a virtual school: Does working memory play a role?. Memory & Cognition, 44(4), 580–589.

Jarratt KP, Riccio CA, & Siekierski BM (2005). Assessment of attention deficit hyperactivity disorder (ADHD) using the BASC and BRIEF. Applied Neuropsychology, 12(2), 83–93.

Joormann J, & Gotlib IH (2008). Updating the contents of working memory in depression: Interference from irrelevant negative material. Journal of Abnormal Psychology, 117(1), 182.

Jovanovski D, Zakzanis K, Campbell Z, Erb S, & Nussbaum D (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test. Applied Neuropsychology: Adult, 19(3), 171–182.

Kamradt JM, Ullsperger JM, & Nikolas MA (2014). Executive function assessment and adult attention-deficit/hyperactivity disorder: Tasks versus ratings on the Barkley Deficits in Executive Functioning Scale. Psychological Assessment, 26(4), 1095.

Kane MJ, & Engle RW (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General, 132(1), 47.

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.

Karr JE, Garcia-Barrera MA, Holdnack JA, & Iverson GL (2018). Advanced clinical interpretation of the Delis-Kaplan Executive Function System: Multivariate base rates of low scores. The Clinical Neuropsychologist, 32(1), 42–53.

Kasper LJ, Alderson RM, & Hudec KL (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. Clinical Psychology Review, 32(7), 605–617.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, & 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 and Adolescent Psychiatry, 36, 980–988.

Kaufman AS, & Kaufman NL (2004). Kaufman Test of Educational Achievement (2nd ed.). Minneapolis, MN: NCS Pearson.

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

Kieffer MJ, Vukovic RK, & Berry D (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. Reading Research Quarterly, 48(4), 333–348.

Kofler MJ, Irwin LN, Soto EF, Groves NB, Harmon SL, & Sarver DE (2019). Executive functioning heterogeneity in pediatric ADHD. Journal of Abnormal Child Psychology, 47(2), 273–286.

Kofler MJ, Rapport MD, Bolden J, Sarver DE, & Raiker JS (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. Journal of Abnormal Child Psychology, 38(2), 149–161.

Kofler MJ, Sarver DE, Harmon SL, Moltisanti A, Aduen PA, Soto EF, & Ferretti N (2018). Working memory and organizational skills problems in ADHD. Journal of Child Psychology and Psychiatry, 59(1), 57–67.

Kofler MJ, Sarver DE, & Wells EL (2015). Working memory and increased activity level (hyperactivity) in ADHD: Experimental evidence for a functional relation. Journal of Attention Disorders.

Kuriyan AB, Pelham WE Jr., Molina BSG, Waschbusch DA, Gnagy EM, Sibley MH, … Kent KM (2013). Young adult educational and vocational outcomes of children diagnosed with ADHD. Journal of Abnormal Child Psychology, 41(1), 27–41.

Kamphaus RW, & Reynolds CR (2007). BASC-2 behavioral and emotional screening system manual. Circle Pines, MN: Pearson.

Lerner MD, & Lonigan CJ (2014). Executive function among preschool children: Unitary versus distinct abilities. Journal of Psychopathology and Behavioral Assessment, 36(4), 626–639.

Lewis C, & Carpendale JI (2009). Introduction: Links between social interaction and executive function. New Directions for Child and Adolescent Development, 2009(123), 1–15.

Lijffijt M, Kenemans JL, Verbaten MN, & van Engeland H (2005). A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: Deficient inhibitory motor control?. Journal of Abnormal Psychology, 114(2), 216.

Mahone EM, Cirino PT, Cutting LE, Cerrone PM, Hagelthorn KM, Hiemenz JR, … & Denckla MB (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. Archives of Clinical Neuropsychology, 17(7), 643–662.

Mahone EM, & Hoffman J (2007). Behavior ratings of executive function among preschoolers with ADHD. The Clinical Neuropsychologist, 21(4), 569–586.

McAuley T, Chen S, Goos L, Schachar R, & Crosbie J (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function?. Journal of the International Neuropsychological Society, 16(3), 495–505.

McInnes A, Humphries T, Hogg-Johnson S, & Tannock R (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. Journal of Abnormal Child Psychology, 31(4), 427–443.

Mesholam-Gately RI, Giuliano AJ, Goff KP, Faraone SV, & Seidman LJ (2009). Neurocognition in first-episode schizophrenia: A meta-analytic review. Neuropsychology, 23(3), 315.

Miller M, & Hinshaw SP (2010). Does childhood executive function predict adolescent functional outcomes in girls with ADHD?. Journal of Abnormal Child Psychology, 38(3), 315–326.

Miller M, Nevado-Montenegro AJ, & Hinshaw SP (2012). Childhood executive function continues to predict outcomes in young adult females with and without childhood-diagnosed ADHD. Journal of Abnormal Child Psychology, 40(5), 657–668.

Miranda A, Colomer C, Mercader J, Fernández MI, & Presentación MJ (2015). Performance-based tests versus behavioral ratings in the assessment of executive functioning in preschoolers: associations with ADHD symptoms and reading achievement. Frontiers in Psychology, 6, 545.

Miyake A, & Friedman NP (2012). The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science, 21(1), 8–14.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, & Wager TD (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49–100.

Moffitt TE (2012). Childhood self-control predicts adult health, wealth, and crime. Multi-Discipl. Symp. Improv. Well-Being Children Youth, Copenhagen.

Morgan AB, & Lilienfeld SO (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. Clinical Psychology Review, 20(1), 113–136.

Muris P, van der Pennen E, Sigmond R, & Mayer B (2008). Symptoms of anxiety, depression, and aggression in non-clinical children: Relationships with self-report and performance-based measures of attention and effortful control. Child Psychiatry and Human Development, 39(4), 455–467.

Muthén BO, & Muthén LK (2012). Mplus 7 base program. Los Angeles, CA: Muthén & Muthén.

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.

Nee DE, & Jonides J (2013). Neural evidence for a 3-state model of visual short-term memory. Neuroimage, 74, 1–11.

Nordvall O, Jonsson B, & Neely AS (2017). Self-reported and performance-based measures of executive functions in interned youth. Psychology, Crime & Law, 23(3), 240–253.

O'Shaughnessy TE, Lane KL, Gresham FM, & Beebe-Frankenberger ME (2003). Children placed at risk for learning and behavioral difficulties: Implementing a school-wide system of early identification and intervention. Remedial and Special Education, 24(1), 27–35.

Parsons TD, Carlew AR, Magtoto J, & Stonecipher K (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. Neuropsychological rehabilitation, 27(5), 777–807.

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.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/.

Raghubar KP, Barnes MA, & Hecht SA (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. Learning and Individual Differences, 20(2), 110–122.

Raiker JS, Rapport MD, Kofler MJ, & Sarver DE (2012). Objectively-measured impulsivity and attention-deficit/hyperactivity disorder (ADHD): Testing competing predictions from the working memory and behavioral inhibition models of ADHD. Journal of Abnormal Child Psychology, 40(5), 699–713.

Rapport MD, Bolden J, Kofler MJ, Sarver DE, Raiker JS, & Alderson RM (2009). Hyperactivity in boys with ADHD: A ubiquitous core symptom or manifestation of working memory deficits? Journal of Abnormal Child Psychology, 37, 521–534.

Renison B, Ponsford J, Testa R, Richardson B, & Brownfield K (2012). The ecological and construct validity of a newly developed measure of executive function: the virtual library task. Journal of the International Neuropsychological Society, 18(3), 440–450.

Reynolds CR, & Kamphaus RW (2004). BASC-2: Behavior Assessment System for Children: Second Edition, manual. Bloomington, MN: Pearson Assessments.

Reynolds CR, & Kamphaus RW (2015). BASC-3: Behavior Assessment System for Children. Bloomington, MN: Pearson.

Samuels WE, Tournaki N, Blackman S, & Zilinski C (2016). Executive functioning predicts academic achievement in middle school: A four-year longitudinal study. The Journal of Educational Research, 109(5), 478–490.

Sarver DE, Rapport MD, Kofler MJ, Raiker JS, & Friedman LM (2015). Hyperactivity in ADHD: Impairing deficit or compensatory behavior?. Journal of Abnormal Child Psychology, 43(7), 1219–1232.

Schmiedek F, Lövdén M, & Lindenberger U (2014). A task is a task is a task: Putting complex span, n-back, and other working memory indicators in psychometric context. Frontiers in Psychology, 5, 1475.

Sesma HW, Mahone EM, Levine T, Eason SH, & Cutting LE (2009). The contribution of executive skills to reading comprehension. Child Neuropsychology, 15(3), 232–246.

Schafer JL (1997). Analysis of incomplete multivariate data. London: Chapman.

Shelton JT, Elliott EM, Matthews RA, Hill BD, & Gouvier WM (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3), 813.

Snyder HR (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. Psychological Bulletin, 139(1), 81.

Snyder HR, Miyake A, & Hankin BL (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. Frontiers in Psychology, 6, 328.

Soreni N, Crosbie J, Ickowicz A, & Schachar R (2009). Stop signal and Conners' CPT: Test-retest reliability of two inhibition measures in ADHD children. Journal of Attention Disorders, 13, 137–43.

Spiegel JA, & Lonigan CJ (2018). A head-to-toes approach to computerized testing of executive functioning in young children. Early Childhood Research Quarterly, 44, 15–23.

Spiegel JA, Lonigan CJ, & Phillips BM (2017). Factor structure and utility of the Behavior Rating Inventory of Executive Function—Preschool Version. Psychological Assessment, 29(2), 172.

St. Clair-Thompson HL, & Gathercole SE (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. Quarterly Journal of Experimental Psychology, 59, 745–759.

Sullivan JR, & Riccio CA (2006). An empirical analysis of the BASC Frontal Lobe/Executive Control scale with a clinical sample. Archives of Clinical Neuropsychology, 21(5), 495–501.

Swanson L, & Kim K (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. Intelligence, 35, 151–168.

Tabachnick BG, & Fidell LS (2013). Using multivariate statistics, 6th edn Boston. Ma: Pearson.

Tarle SJ, Alderson RM, Patros CH, Arrington EF, & Roberts DK (2018). Working memory and behavioral inhibition in children with attention-deficit/hyperactivity disorder (ADHD): An examination of varied central executive demands, construct overlap, and task impurity. Child Neuropsychology, 25(5), 664–687.

Thorell LB (2007). Do delay aversion and executive function deficits make distinct contributions to the functional impact of ADHD symptoms? A study of early academic skill deficits. Journal of Child Psychology and Psychiatry, 48(11), 1061–1070.

Toplak ME, Bucciarelli SM, Jain U, & Tannock R (2008). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). Child Neuropsychology, 15(1), 53–72.

Toplak ME, West RF, & Stanovich KE (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct?. Journal of Child Psychology and Psychiatry, 54(2), 131–143.

Unsworth N, & Engle RW (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. Psychological Review, 114(1), 104.

Unsworth N, Spillers GJ, & Brewer GA (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(1), 240.

van der Ven SH, Kroesbergen EH, Boom J, & Leseman PP (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. British Journal of Developmental Psychology, 31(1), 70–87.

Vriezen ER, & Pigott SE (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. Child Neuropsychology, 8(4), 296–303.

Wager TD, & Smith EE (2003). Neuroimaging studies of working memory. Cognitive, Affective, & Behavioral Neuroscience, 3(4), 255–274.

Wåhlstedt C, Thorell LB, & Bohlin G (2009). Heterogeneity in ADHD: Neuropsychological pathways, comorbidity and symptom domains. Journal of Abnormal Child Psychology, 37(4), 551–564.

Walker A, & Nabuzoka D (2007). Academic achievement and social functioning of children with and without learning difficulties. Educational Psychology, 27(5), 635–654.

Wechsler D (2014). Wechsler Intelligence Scale for Children-Fifth Edition. San Antonio, TX: The Psychological Corporation.

Wechsler D (2011). Wechsler Abbreviated Scale of Intelligence-Second Edition. San Antonio, TX: The Psychological Corporation.

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.

Wentzel KR (1991). Relations between social competence and academic achievement in early adolescence. Child Development, 62(5), 1066–1078.

Wiebe SA, Sheffield T, Nelson JM, Clark CA, Chevalier N, & Espy KA (2011). The structure of executive function in 3-year-olds. Journal of Experimental Child Psychology, 108(3), 436–452.

Willcutt EG, Doyle AE, Nigg JT, Faraone SV, & Pennington BF (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. Biological Psychiatry, 57(11), 1336–1346.

Willoughby MT, & Blair CB (2016). Measuring executive function in early childhood: A case for formative measurement. Psychological Assessment, 28(3), 319.

Key Points

Question: Executive functions provide an important foundation for children's behavior and learning, but what method is best for assessing these abilities in children?

Findings: Executive function tests and ratings share less than 10% of their variance at the construct level, and as such cannot be used interchangeably to measure children's executive functions.

Importance: Executive function tests showed stronger evidence of validity for predicting academic functioning. Given the strong theoretical and empirical links between executive functioning and academics, caution is recommended when drawing conclusions about children's neurocognitive functioning based on rating scales.

Next Steps: Executive function rating scales provide a reliable method for assessing children's behavior. However, the behaviors captured by these rating scales appear to be related to brain-based executive functioning abilities to only a small degree, which raises questions about whether the same term should be used to describe the underlying neurocognitive abilities and the presumed behavioral outcomes of those abilities. Given that these behaviors uniquely predict children's academic attainment, future work is needed to identify the neurocognitive mechanisms and processes underlying these 'executive' behaviors, toward clarifying the construct space of executive functions across different research traditions.
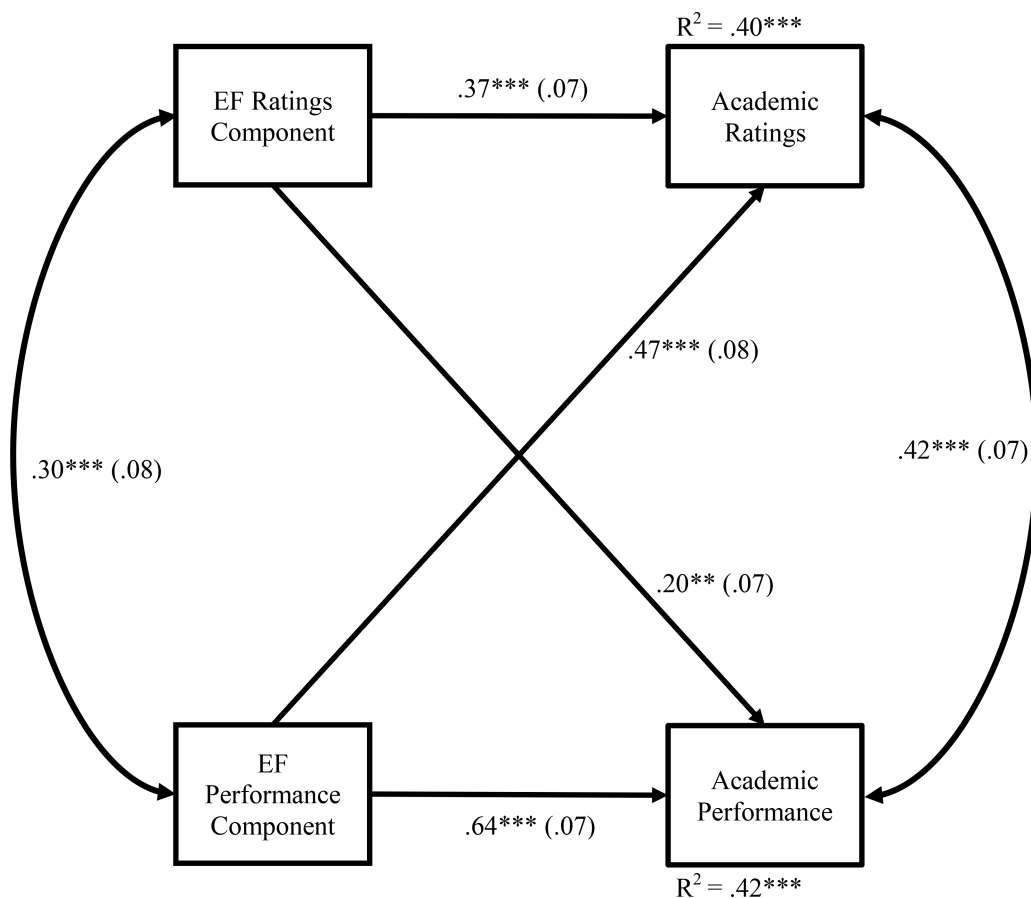
Figure 1.
Executive function ratings and test performance predicting academic ratings and test performance. <u>Standardized</u> path coefficients are shown. Age, sex, medication status (no/yes), and ADHD status (no/yes) were covaried but not shown for clarity. Detailed parameter estimates are given in Supplementary Tables S2 and Table S3. EF = Executive Function.
Note: * *p* < .05, ** *p* < .01, *** *p* < .001

Table 1.

Demographics

Variable	Sample		Range	Skew	Kurtosis
	M	**SD**			
N (Boys/Girls)	136 (88/48)		--	--	--
Age	10.34	1.50	5.04	0.47	−1.07
SES	48.48	11.54	46.00	−0.55	−0.40
VCI	105.62	12.65	58.00	0.12	−0.37
Ethnicity (A, BR, B, H, W)	(3, 9, 15, 17, 92)		--	--	--
Executive Function Performance Measures					
PH Working Memory					
Set Size 3	2.76	0.29	1.12	−1.21	0.52
Set Size 4	3.41	0.57	2.60	−1.00	0.45
Set Size 5	3.96	0.81	3.33	−0.64	−0.45
Set Size 6	3.60	1.22	5.67	−0.24	−0.52
VS Working Memory					
Set Size 3	2.15	0.64	2.83	−0.99	0.40
Set Size 4	2.69	1.00	3.83	−0.61	−0.57
Set Size 5	2.81	1.11	4.33	−0.14	−0.88
Set Size 6	2.63	1.20	5.33	0.49	−0.28
Stop Signal					
Block 1 (SSD)	281.07	74.86	331.25	−0.16	−0.34
Block 2 (SSD)	277.71	87.49	331.25	−0.32	−0.88
Block 3 (SSD)	277.62	83.66	331.25	−0.28	−0.62
Block 4 (SSD)	287.87	85.15	331.25	−0.58	−0.33
Go/No-Go					
Block 1 (Commission Errors)	0.40	0.62	2.00	1.33	0.65
Block 2 (Commission Errors)	0.50	0.70	2.00	1.06	−0.20
Block 3 (Commission Errors)	0.74	0.90	3.00	1.03	0.13
Block 4 (Commission Errors)	1.10	1.19	5.00	0.98	0.36
Executive Function Ratings Measures					
BASC-2/3: Teacher GEC (T-score)	58.66	11.39	51.00	0.33	−0.28
BASC-2/3: Parent GEC (T-scores)	62.26	11.02	47.00	−0.07	−0.61
BRIEF: Teacher Executive Function (T-score)	68.39	15.55	72.00	−0.26	−0.66
BRIEF: Parent Executive Function (T-score)	64.56	11.19	51.00	0.26	−0.54
Academic Achievement Measures					
KTEA: Academic Skills Battery (Composite scores)	100.31	12.94	64.50	0.35	0.12
APRS: Academic Productivity (T-score)	44.66	9.58	44.00	−0.01	−0.18

Note: A = Asian; B = Black/African American; BR = Biracial; H = Hispanic/English-speaking; PH = Phonological; SES = Hollingshead SES total score; VCI = Verbal Comprehension Index; VS = Visuospatial; W = White/Non-Hispanic

Table 2.

Associations among study variables

	Academic Performance	Academic Ratings	Executive Function Tests	Executive Function Ratings	Age	Sex	Meds (N/Y)
1. Academic Performance	--						
2. Academic Ratings	.63 ***	--					
3. EF Tests Component	.55 ***	.46 ***	--				
4. EF Ratings Component	.36 ***	.50 ***	.30 ***	--			
9. Age	.03, *ns*	-.02, *ns*	.47 ***	.04, *ns*	--		
10. Sex (Female/Male)	.06, *ns*	.10, *ns*	.10, *ns*	.19 *	.08, *ns*	--	
11. Medication Status (No/Yes)	.06, *ns*	-.07, *ns*	.04, *ns*	-.06, *ns*	.24 **	.23 **	--
12. ADHD Status (No/Yes)	-.34 ***	-.37 ***	-.63 ***	-.31 ***	-.18 *	.14, *ns*	.15, *ns*

Note:

*
= *p*<.05

**
= *p*<.01

***
= *p*<.001

ns = non-significant (*p*>.05).

Correlations in boxes reflect associations between different measures intended to measure the same construct. Acad = Academic; EF = Executive Function; IC = Inhibitory Control; Meds = Medication Status (No = 0, 1 = Yes); Sex (Female = 0, Male = 1); WM = Working Memory