Published in final edited form as:

J Psychopathol Behav Assess. 2014 December ; 36(4): 626–639. doi:10.1007/s10862-014-9424-3.

Executive Function Among Preschool Children: Unitary Versus Distinct Abilities

Matthew D. Lerner[1] and **Christopher J. Lonigan**[1,2]

[1]Department of Psychology, Florida State University

[2]Florida Center for Reading Research, Florida State University

Abstract

Working memory (WM) and inhibitory control (IC) are considered related but separable executive functions (EFs) among adults and adolescents. Although available evidence suggests that these constructs have not diverged especially among younger preschool children, questions remain regarding the age at which separable factors emerge. This study used confirmatory factor analysis to test a 2-factor model of EF among 289 preschool children whose ages ranged from 45 to 63 months (M = 55.74, SD = 7.56). As hypothesized, the model including separate but related factors provided a significantly better fit than a unitary model, indicating the presence of distinct WM and IC factors. Based on evidence that WM and IC measured during preschool relate differently to a variety of academic and behavioral outcomes, it was hypothesized that a model including separate factors for each EF would fit the observed data better than a single-factor model. Although the two-factor model provided the best fit for the full sample, the correlation between WM and IC factors was significantly higher for younger (φ =.95) than older (φ =.68) children, indicating increasing divergence as a function of age.

Keywords

executive function; inhibitory control; working memory; preschool children

Executive functions (EFs) are the fundamental capacities that underlie purposeful use of attention and make goal-driven behaviors possible. The most frequently studied EFs among adults are (a) working memory (WM), the updating or active use of information held in memory, (b) inhibitory control (IC), the suppression of a predisposed or learned response, and (c) shifting (SH), the alternation between sets of stimulus-response rules. Findings from studies using confirmatory factor analysis (CFA) with data from adult samples have demonstrated that EF tasks are best represented as three distinct but interrelated domains (i.e., WM, IC, SH) among adults. For example, Miyake, Friedman, Emerson, Witzki, Howerter, and Wager (2000) reported that a three-factor oblique model of EF provided a better fit to data from adults than did a single-factor model, any two-factor model, or a three-factor orthogonal model.

Questions regarding this work should be sent to Matthew D. Lerner (lerner@psy.fsu.edu; phone: 706-424-2254; fax 850-644-7739) or Christopher J. Lonigan (lonigan@psy.fsu.edu; phone: 850-644-7241; fax: 850-644-7739), Department of Psychology, Florida State University, 1107 W. Call St., Tallahassee FL 32306-4301.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In the past decade, an increasing emphasis has been placed on understanding the nature, development, and correlates of EF among progressively younger populations. In a CFA study of 9- to 13-year-old children, EF was best represented by the same factor structure observed among adults, namely, interrelated but separable factors representing WM, SH, and IC (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003). Similarly, Shing, Lindenberger, Diamond, Li, & Davidson (2010) reported separable EF components (memory maintenance and IC) among 10- to 15-year-old children but a unitary structure among children ages 4–7 and 7–9.5 years. In contrast, prior to school entry, evidence regarding the dimensionality of EFs is less clear, with some studies indicating a unitary factor (Willoughby, Blair, Wirth, & Greenberg, 2010; Wiebe, Sheffield, Nelson, Clark, Chevalier, & Espy, 2011) and others indicating that WM and IC relate differently to a range of outcomes, including concurrent and subsequent performance in mathematics and reading (e.g., Bull, Espy, & Wiebe, 2008; Bull & Scerif, 2001), as well as internalizing (e.g., Rhoades, Greenberg, & Domitrovitch, 2009), and externalizing (e.g., Brocki, Eninger, Thorell, & Bohlin, 2010) behavior problems.

EF first becomes evident during the first year of life (Johnson, 2005; Pelphrey et al., 2004; Reznick, Morrow, Goldman, & Snyder, 2004) and undergoes rapid development between three and six years of age (e.g., Carlson & Moses, 2001). Therefore, understanding the nature of EF prior to school entry is particularly important. Although rudimentary forms of WM have been measured among children younger than 36 months (Corrigan, 1981), the extent to which these tasks align with the definition of WM used with adults and older children (i.e., manipulation or updating of information held in memory) is unclear. Given the absence of evidence that WM, as defined in studies of older children and adults, can be measured reliably among children younger than three years of age, this is likely the earliest age at which the divergence between WM and other component EFs can be tested. A clearer understanding of the dimensionality of EF immediately prior to formal school entry (i.e., among older preschool children) might provide increased precision in identifying potential causal mechanisms responsible for the acquisition of academic and socioemotional skills.

Factor Analytic Studies of EFs in Young Children

Analytic (i.e., inability to test competing models) or methodological (i.e., task selection, sample characteristics) factors limit the extent to which each of the six previous factor analytic studies can determine the age at which EF components first diverge. Two studies employed exploratory factor analysis (EFA; Hughes & Ensor, 2007; Welsh, Nix, Blair, Bierman, & Nelson, 2010) and each reported that the underlying structure of EF was adequately represented by a single factor. As stated by Welsh et al. (2010) EFA “does not shed light on the important conceptual and measurement/issues regarding executive functions” (p. 50) because such models cannot determine the degree of improvement in model fit that might result from the inclusion of an additional factor. This distinction is particularly important when a significant correlation is expected between the potentially distinct factors, as is the case with EF components.

With regard to task selection, Willoughby et al. (2010) used a single task to index WM, which prevents an examination of the boundary between WM and other constructs because the extent to which performance on a single task is determined by task-specific features as opposed to the underlying construct cannot be determined. The same study tested a two-factor model that included an IC factor composed of three IC tasks and a second factor composed of one WM task and one SH task. A latent variable composed of two incongruous tasks would not be expected to reliably index either construct, especially if there were a reason to expect a negative relation between those two constructs. WM and SH have been shown to correlate negatively among young children, perhaps because systematic perseverative errors cannot occur unless a previous rule is held in memory (Jones, Rothbart, & Posner, 2003).

In other cases, the relations between established construct definitions and selected tasks have been unclear, resulting in the combination of tasks that vary in the extent to which they fit the definition of the intended EF component. For example, the act of manipulating or updating information held in memory has been demonstrated as the essential component of WM (Gathercole & Pickering, 2000; Miyake et al., 2000); however, the group of tasks intended to measure WM in the Wiebe et al. (2008) sample includes a simple forward span task (Recall for Digits; DAS-I, Elliott, 1990), which requires recall but neither updating nor manipulation. Thus, the failure of these three memory tasks to form a coherent factor could be due to the combination of WM tasks with a simple forward span task. Similar questions arise with regard to the IC tasks in the same study. Visual Attention (NEPSY; Korkman, 1998) is, at least in significant part, an index of response speed, especially given that the dependent variable was calculated by dividing the number of correct responses by the response time. Similarly, the inhibit condition of the Shape School task (Espy, 1997) includes a two-part if-then response rule, bringing the demands of this task closer to the definition of SH than IC. Finally, the CPT (Kerns & Rondeau, 1998) was validated as a measure of sustained attention, not IC, and differs importantly from typical IC tasks in that it requires five minutes of continuous performance of a task that would not otherwise be difficult (i.e., discriminating between two types of animals). When one or more of the factors used in a model comparison includes measures typically used to assess other constructs, the failure of more complex models to improve model fit is not surprising, but its meaning with regard to the WM-IC boundary remains unclear.

With regard to sample characteristics, stronger evidence regarding the dimensionality of EF using a CFA framework and well-defined tasks is provided by two studies with highly specific samples. Wiebe et al. (2011) provides clear evidence of the absence of a distinction between WM and IC at age three (M = 3.01 years, *SD* = 12.8 days) but was not designed to address the question of divergence between component EFs in a larger age range. Schoemaker et al. (2012), reported separate but correlated WM and IC factors in a sample of three-to-five-year-old children, each of whom had been diagnosed with one or more externalizing disorders. Thus, a similarly-designed study with a community sample of preschool children slightly older than those examined by Weibe et al. would build on both the finding of unidimensional EF at age three and that of separable components in a clinical sample by seeking to determine the youngest age at which separable factors can be measured in a typically developing population.

Academic and Behavioral Correlates of EFs

In contrast to the majority of results from factor analytic studies, results of concurrent and longitudinal correlational studies typically identify component-specific predictive relations between EF and important developmental outcomes. For example, IC is positively associated with concurrent internalizing symptoms and negatively associated with externalizing symptoms among 6-to-12-year-old children (Kooijmans, Scheres, & Oosterlaan, 2000). IC in first grade predicts higher social competence and fewer externalizing symptoms in third grade, even after controlling for initial levels of the outcome variables and IC measured in the third grade (Nigg, Quamma, Greenberg, & Kusche, 1999). More recent evidence indicates that IC in particular may be responsible for the influence of EF on change in internalizing and externalizing symptoms during early grade school years (Riggs, Blair, & Greenberg, 2003). Of particular relevance to determining the dimensionality of EF, when multiple components of EF are included as predictors, the relation between EF and behavior appears to be specific to IC (e.g., Brocki, Eninger, Thorell, & Bohlin, 2010). EFs also appear to relate closely and independently to academic outcomes among grade-school children. For example, second grade WM predicts third grade reading comprehension, after controlling for second grade reading comprehension, vocabulary, and nonword reading (Seigneuric & Ehrlich, 2005). Alloway, Gathercole, Kirkwood, and Elliott (2009) replicated this relation and expanded it to include both reading and mathematics, demonstrating that WM predicted mathematics and reading performance among both six- and eleven-year-old children, after controlling for IQ and receptive vocabulary.

Similar to evidence regarding grade-school children, EFs are significantly associated with internalizing and externalizing symptoms, adaptive classroom behaviors, and academic achievement among preschool children. Hughes, White, Sharpen, and Dunn (2000) demonstrated concurrent relations between complex EF tasks (i.e., tasks known to tap multiple EFs among adults) and observer-rated social behavior among four-year-old children; these relations were independent of verbal IQ and maternal education. Raaijmakers et al. (2008) compared four-year-old children who scored at or above the 93rd percentile on the aggressive behavior subscale of the Children’s Behavior Checklist (CBCL; Achenbach & Rescorla, 2000) with children from the same sample, matched for IQ, who scored below the 50th percentile on the same aggression scale. High scores for aggression were significantly related to the IC cluster of tasks but not to tasks intended to measure other areas of EF, suggesting that the relation between EF and behavioral outcomes may be specific to IC as early as age four years of age. Bull et al. (2008) reported that preschool nonverbal WM significantly predicted math but not reading performance at age seven, whereas preschool IC significantly predicted age seven performance in reading but not math, again suggesting that when multiple component EFs are measured during preschool, divergent predictive patterns emerge.

Although the nature of EF is fairly well understood among adults, less is known about this fundamental capacity in the years during which it develops most rapidly. EF has consistently been shown to relate closely to academic and behavioral outcomes, and when multiple component EFs have been included simultaneously, patterns of differential prediction emerge as early as preschool. However, findings from studies that have directly examined the dimensionality of EF among preschool children seem to indicate the absence of a distinction between WM and IC during preschool, especially with regard to non-clinical samples. Therefore, open questions remain regarding the extent to which the apparent absence of a boundary between WM and IC during preschool is the result of: (a) increasing distinction between WM and IC throughout the preschool years, such that the boundary is present among older preschoolers but obscured by examining data from younger and older children simultaneously, (b) the use of measurement and analytic procedures that do not maximize the probability of detecting a boundary if one were present, or (c) the actual absence of a boundary between the constructs of WM and IC during preschool.

Current Study

The purpose of this study was to examine the dimensionality of WM and IC among preschool children and to detect any age- or gender-related differences in the relation between WM and IC. Multiple tasks hypothesized to represent each EF component were used, and competing models were examined to determine if a multi-factor EF model provided a significantly better representation of child performance than a unitary model and if response suppression was distinguishable from response conflict among IC tasks. Multi-group analyses were used to determine if the tasks selected represented the same underlying abilities in both subgroups and if the best-fitting model was equally applicable, when comparing younger versus older children and girls versus boys in the sample. In light of evidence that WM and IC relate differently to a variety of academic and behavioral outcomes among preschool children, it was hypothesized that these two EF components would form separable but correlated constructs among preschool children, as they do among adults and older children, and that this factor structure would be equally applicable to younger and older children in the sample. In the absence of evidence to suggest sex differences in EF structure, the same two-factor model was expected to provide the best fit for girls and boys. Only the question of the distinction between IC and WM was addressed by this study. Tasks hypothesized to represent SH were not included. Evidence linking EF to divergent patterns of academic and behavioral outcomes in preschool and early-grade school has focused primarily on IC and WM. Moreover, only one well-validated SH measure has been widely shown to function adequately with four-year-old children (i.e., Dimension Change Card Sort; Frye, Zelazo, & Palfai, 1995), and at least three measures of SH would have been needed to identify a SH factor.

Method

Participants

The sample included 289 children (53% female) recruited from 19 preschools and child care centers serving a diverse population of children in north Florida (e.g., eight of the preschools were located in neighborhoods served by Title 1 classified elementary schools). These children ranged in age from 45 to 63 months (M = 55.74, *SD* = 7.56). Given the evidence of unidimensional EF among three-year-old children and the goal of determining the age at which component EFs diverge, emphasis was placed on recruiting four-year-olds. The sample was ethnically diverse; it was composed of 31% African American/Black, 57% Caucasian/White, 3% Hispanic/Latino, 3% Asian, and 7% other ethnicities.

Measures

Working memory—Consistent with the Miyake et al. (2000) definition of WM, each task required children to update or actively manipulate information held in memory in response to newly presented stimuli. WM is often tested among adults by requiring strings of letters or numbers to be repeated in a different order than they were presented (e.g., backward, alphabetical). To prevent the confounding of WM with letter or number knowledge, tasks used in this study did not use letters or numbers.

Word span reversed: This task was similar in demands and administration to the digit span reversed subtest of the WISC-IV (Wecshler, 2003) but instead of lists of digits, children were presented with lists of common words and asked to repeat them in reverse order. Word lists ranged in length from two to eight words, with three trials at each level. For example, during the first three-item trial, children were presented with the list "nose, hand, car," and the correct response was "car, hand, nose." The dependent variable was the total number of trials in which all words were repeated in the correct order, resulting in a maximum possible score of 21. Cronbach's alpha for Word Span Reversed was .72 in this sample.

Size ordering: This task was modeled after the Children's Size Ordering Task (McInerney, Hrabok, & Kerns, 2005) using simpler vocabulary commonly found in measures of vocabulary developed for three- and four-year-old children. Children were presented orally with lists of common objects (e.g., car, house, bird) and asked to repeat them in order from the smallest to the largest object (e.g., bird, car, house). Word lists ranged in length from two to eight words, with three trials at each level. The dependent variable was the total number of trials in which all words were repeated in the correct order, resulting in a maximum possible score of 21. Cronbach's alpha for Size Ordering was .61 in this sample.

Object span: On this task, children had to recall the names of objects previously presented as pictures. Children were presented with two cards face down on the table. The first card was turned face up, and children were asked to name the objects displayed on the card before it was turned face down. Next, children were shown the second card and asked to name the objects on it before it was turned face down. Children were then asked to recall the objects on the first card. The number of objects on each card increased from two to eight and was the same for both cards in a trial; there were three trials at each level. The dependent variable was the number of trials in which all objects were correctly named from the first card, resulting in a maximum possible score of 21. Cronbach's alpha for Object Span was .27 in this sample. Due to the poor internal consistency of Object Span, model comparisons were performed with and without this task, but results were unchanged.

Listening span: On this task, children had to recall words presented verbally. Children were asked to respond to increasingly long lists of simple questions (e.g., Do dogs bark? Can cats fly?). After giving a response, they were asked to recall the last words of each question. This task was similar to the listening recall subtest of the Working Memory Test Battery for

Children (Pickering & Gathercole, 2001) but used questions with vocabulary appropriate for preschool children. Questions were asked in sets of two, three, and four, with three trials at each level. The dependent variable was the total number of words from the non-last sentences (i.e., those subject to the updating demand posed by the last sentence) recalled correctly, resulting in a maximum possible score of 60. Cronbach's alpha for Listening Span was .89 in this sample.

Inhibitory control, response suppression—Children completed three IC tasks reflecting situations in which IC was measured with suppression. In suppression tasks, children were required to refrain from performing a dominant response by simply doing nothing.

Bird and Dragon: On this task, adapted from (Reed, Pien, & Rothbart, 1984) and (Kochanska, Murray, Jacques, Koening & Vandegeest, 1996), children completed five practice trials, during which they learned to obey simple commands from the bird (a puppet who speaks in a high voice) and to ignore simple commands from the dragon (another puppet who speaks in a low voice). A total of 25 trials included 12 randomly-ordered critical (dragon) trials in which the child received a score from 0–3 (0 points for full movement, 1 point for wrong movement, 2 for self-correction, 3 for no movement). The dependent variable was the average score across 12 critical trials, resulting in a maximum possible score of 3. Cronbach's alpha for Bird and Dragon was .95 in this sample.

Luria's Hand Game: This task was adapted from Hughes (1996). Children learned in practice trials to imitate the examiner's gesture, either a fist or one finger pointing. To target suppression, as opposed to response conflict, in test trials children were instructed to imitate only the pointing gesture but to do nothing when the examiner showed a fist. Before each trial, children were required to place their dominant hand on a flat surface. A total of 25 trials included 12 randomly-ordered critical (fist) trials, with scores of 0–3 derived identically to Bird and Dragon scoring. The dependent variable was average score across 12 critical trials, resulting in a maximum possible score of three. Cronbach's alpha for Luria's Hand Game was .76 in this sample.

Picture Imitation: During practice trials, children learned to imitate black-and-white line drawings of animals and children shown touching their heads or toes. After the dominant response was established during imitate trials, children were instructed to continue imitating animals but to do nothing when they saw a child (the child picture was the same gender as the participating child). Children were shown 12 cards, one at a time, half of which represented randomly-ordered critical (child image) trials. Before each trial, children were required to place their dominant hand on a flat surface. Scores of 0–3 were derived identically to Bird and Dragon scoring. The dependent variable was average score across six critical trials, resulting in a maximum possible score of 3. Cronbach's alpha for Picture Imitation was .91 in this sample.

Inhibitory control, response conflict—Children completed three IC tasks reflecting situations in which IC was measured with response conflict. In response conflict tasks, children were required to refrain from a dominant response and also perform a sub-dominant response.

Block sorting: Children first learned to sort red and white blocks into red and white bowls, respectively. Bowls were placed upside down on the table, and blocks were inserted through a small opening in the bottom of the bowl. After the dominant response was established, children were instructed to place red blocks in the white bowl and white blocks in the red bowl. Children were given 12 blocks, one at a time, in a randomized order. Children received scores from 0–3 (0 points for a block in the same colored bowl, 1 point for any other incorrect placement, 2 points for a self-correction [e.g., any movement of the hand toward the same-colored bowl, followed by placing the block in the correct bowl, within two seconds, without examiner feedback], and 3 points for a block in the opposite color bowl). The dependent variable was average score across 12 trials, resulting in a maximum possible score of 3. Cronbach's alpha was .83 for Block Sorting in this sample.

Day-night: (Gerstadt, Hong, & Diamond, 1994) Children were shown cards with cartoon drawings, one depicting the sun and one depicting the moon. Examiners first established that the child knew that the sun is seen during the day and the moon is seen during the night. Instead of a verbal response, children were asked to point to the sun card when the examiner said "night" and the moon card when the examiner said "day." a Responses on a total of 12 trials were scored 0–3 (0 for pointing incorrectly, 1 for pointing to anything but the wrong answer, 2 for self-correction [e.g., any movement of the hand toward the incorrect card, followed by indicating the correct card, within two seconds, without examiner feedback], and 3 for pointing correctly). The dependent variable was average score across these 12 trials, resulting in a maximum possible score of 3. Cronbach's alpha was .81 for Day-Night in this sample.

Knock-tap: This task is part of the Developmental Neuropsychological Assessment (NEPSY; Korkman, Kirk, & Kemp, 1998). During practice trials, children learned to imitate the examiner's hand gestures, either knocking on the table with the knuckles or tapping on the table with an open palm. After the dominant response had been established, children were instructed to perform the opposite of the examiner's gesture (i.e., knock when examiner tapped). Responses on a total of 12 trials were scored 0–3 (0 for imitating the examiner, 1 for performing some other gesture, 2 for self-correction [e.g., performing an incorrect gesture and then responding correctly, within two seconds, without examiner feedback], and 3 for a correct response), and the dependent variable was average score across these 12 trials, resulting in a maximum possible score of 3. Cronbach's alpha was .90 for Knock-Tap in this sample.

Oral language—Children's oral language skills were assessed using the Oral Language/ Vocabulary subtest of the Voluntary Pre-Kindergarten Screener (VPK screener; Author Omitted). This 23-item task measures children's expressive and receptive vocabulary, awareness of bound morphemes (e.g., "-ed" for past tense or "-s" as a plural marker), and the understanding of adjectives and prepositions (e.g., "point to the line that is *straight*," "point to the cat that is *inside* the box"). Before administering the test items, two practice items were administered, with feedback given for correct and incorrect answers. Test items were administered without corrective feedback, and each child was administered all items. Cronbach's alpha for this subtest was .99 in this sample.

Procedures

Informed consent was obtained from children's parents or guardians before data collection began, and approval was obtained from the Institutional Review Board of the Florida State University. Assessments of children were completed in a quiet area within each child's preschool. Assessments were administered either by clinical psychology graduate students or by undergraduates who were pursuing a degree in psychology and were supervised by a clinical psychology graduate student. Before administering assessments, all examiners were trained and then tested until they could administer all tasks without error. To prevent systematic order effects, two randomly determined orders were created, and children were randomly assigned to receive the tasks in one order or the other. Within each task order, three testing sessions were created, with the restriction that both IC and WM tasks were included in all three testing sessions. For example, in Order 1, Session A included Knock-Tap, Word Span Reversed, Luria's Hand Game, and Block Sorting; Session B included Picture Imitation, Listening Span, and Bird Dragon; and Session C included Object Span, Size Ordering, and Day-Night. Tasks were administered on three separate days during 30- to 45-minute sessions. During these sessions, children were given breaks upon request or if fatigue became apparent to the examiner. Younger and older children received exactly the same tasks with exactly the same rules, stimuli, and instructions. For all 10 EF tasks, corrective feedback was provided during practice trials, but not during any testing trials.

Results

Descriptive Statistics and Preliminary Analyses

Descriptive statistics and zero-order correlations for the 10 EF tasks are shown in Table 1. Between 14 and 19 children had missing data on each task. Correlations between pairs of tasks were generally modest and significant. On two tasks in each group (Word Span Reversed and Size Ordering; Bird-Dragon and Picture Imitation) older children scored significantly higher than younger children, as shown in Table 3. The average correlation between IC tasks (r = .31) was greater than the average correlation between WM tasks (r = .23) and the average correlation between IC and WM tasks (r = .22). Inspection of the distributional qualities of the EF measures indicated that scores on WM tasks were normally distributed, but some of the IC tasks were moderately skewed, kurtotic, or both. Simulation studies in structural equation modeling have indicated that skew values greater than 3.0 (Curran, West, & Finch, 1996) and kurtosis values greater than 10.0 (Kline, 2011) can result in model misspecification. None of the IC variables had skew (range: −2.73 to −1.06; see Table 1) or kurtosis (range: −6.09 to .03) values this extreme; therefore, the observed distributions of the IC variables would not be expected to result in model misspecification.

Dimensionality of Executive Functions

Fit statistics for five models were compared to determine the best representation of the WM and IC components of EF. Model 1 was a one-factor model in which all 10 tasks defined a single factor. Model 2 was a two-factor model in which the six IC tasks (i.e., Bird & Dragon, Luria's Hand Game, Block Sorting, Day-Night Stroop, Knock-Tap, Picture Imitation) defined an IC factor, and the four WM tasks (i.e., Word Span Reversed, Listening Span, Size Ordering, Object Span[1]) defined a WM factor. Model 3 was a two-factor model in which the three IC-suppression tasks (i.e., Bird & Dragon, Luria's Hand Game, Picture Imitation) defined an IC-suppression factor and the four WM tasks along with the three IC-conflict tasks (i.e., Day-Night, Knock-Tap, Block Sorting), defined a combined IC-conflict/WM factor. Model 4 was a two-factor model in which the three IC-conflict tasks defined an IC-conflict factor and the four WM tasks along with the three IC-suppression tasks, together defined a combined IC-suppression/WM factor. Model 5 was a three-factor model in which the WM tasks, IC-suppression tasks, and IC-conflict tasks each defined a separate factor. Across Models 2 – 5, WM and IC factors were allowed to correlate.

CFAs were performed with Mplus 6.1 (Muthén & Muthén, 2011) using full information maximum likelihood to account for missing data. The Yuan-Bentler scaled chi-square (Y-B χ^2; Yuan & Bentler, 1996) was used to correct standard errors for non-normal distributions and as an index of overall model fit. Non-significant values for the Y-B χ^2 statistic represent the most stringent test of absolute model fit. Models with comparative fit index (CFI) and Tucker-Lewis index (TLI) greater than or equal to .95 and root mean squared error of approximation (RMSEA) less than or equal to .06 are considered to fit the data well (Hu & Bentler, 1999). The χ^2 difference test was used to evaluate the significance of any improvement in model fit provided by the inclusion of additional factors.

CFAs were conducted with and without a sandwich estimator to provide adjustments to standard errors and model fit statistics because children were nested in preschools. Both sets of analyses yielded identical results for model comparisons and substantively identical results for multi-group tests. Therefore, results are reported using the standard maximum likelihood (ML) estimator, for which clearer standards exist by which to judge indices of model fit. Preliminary analyses indicated that allowing the residuals of the Bird Dragon and the Picture Imitation tasks to correlate resulted in improved model fit. Although including this correlation in the model improved overall model fit statistics, the results of the model comparison were the same with and without it.

Indices of absolute and relative model fit, shown in Table 2, indicated that Model 1 provided reasonable fit to the data, and Model 2 provided significant improvement compared to Model 1. Despite demonstrating acceptable absolute model fit, neither of the other two-factor models (i.e., Models 3 & 4) provided better fit than the one-factor model, and the three-factor model did not provide improved model fit relative to the two-factor model with separate IC and WM factors (i.e., Model 2). In the two-factor model, the IC factor and the WM factor were significantly correlated ($\varphi = .78$; $p < .001$). Each task loaded significantly

[1]Because OS had the lowest factor loading among the four WM tasks (.34, $p < .001$) and the WM factor accounted for less of the variance in this task ($r^2 = .11$, $p < .01$) than the other three WM tasks (loadings ranged from .53–.60, r^2s's ranged from .30–.36, all *p*s < .001), model comparisons were conducted with and without this task. The OS task was included in the final model because (a) the inclusion of this task did not change the results of the model comparison, (b) the factor loading and variance accounted for were statistically significant, and (c) removing any tasks included in the original model would make the model less strictly confirmatory.

on its respective factor (*p*s < .01; see Table 3), and the factors accounted for between 11 and 55 percent of the variance in the individual tasks.

All of the WM tasks required a verbal response, but all of the IC tasks required a nonverbal response, providing a potential alternative explanation for the better fit of the two-factor model (i.e., response requirement vs. construct measured). To evaluate the extent to which differences in children's oral language skills--and, hence, their abilities to produce the required verbal response readily--might account for the divergence between WM and IC factors, the same series of model comparisons was conducted while statistically controlling for oral language performance on each task. Scores on each EF measure were regressed on scores from the VPK Oral Language/Vocabulary subtest (*r*s ranged from .11–.51; *M* = .34, *SD* = .14), and the standardized residuals were used in the CFAs. Results of these analyses are shown in the lower panel of Table 3. As with the CFA of raw scores on the EF measures, Model 2 provided the best fit and represented a significant improvement over Model 1.

Structural and Measurement Invariance

Multi-sample models were used to evaluate possible differences in how WM and IC were measured and in the degree of divergence between these components when comparing younger versus older children and girls versus boys. Given that the vast majority (82.4%) of the present sample was between four and five years of age, a comparison between four- and five-year-olds (although certainly of theoretical interest) would have had insufficient power to detect age-related differences in EF dimensionality. To maximize the power to detect age-related differences in the present sample, a median split was used to create older and younger groups of equal size. Significant differences in the loadings of tasks on a given factor would mean that the tasks did not represent that factor equivalently across groups. Significant differences in the correlation between factors would mean the degree of overlap between the WM and IC factors was larger in one group than the other. A multi-sample model with none of the parameters constrained to equality across groups served as the basis for comparing the effects of constraining parameters across groups to equality. For both sets of multi-group analyses, increasingly stringent sets of constraints were imposed and included: (a) factor loadings, (b) correlations between the IC and WM factors and between the residuals of the Bird Dragon and the Picture Imitation tasks, and (c) the residuals of each observed variable. Finally, to directly test for possible group differences with regard to the central hypothesis (i.e., the presence of a boundary between WM and IC), the fully unconstrained model was compared to a model in which only the correlation between the WM and IC factors was constrained to equality.

Multi-sample models for younger versus older children—Descriptive statistics for task performance for the younger (M = 52.62 *SD* = 2.68 months) and older (M = 59.28 *SD* = 2.30 months) children in the sample are shown in Table 3. As seen in the table, younger children had significantly lower scores than did older children on two WM tasks (Word Span Reversed, Size Ordering) and two IC tasks (Bird & Dragon, Picture Imitation). A summary of the multi-sample analyses for the younger versus older children is shown in the upper panel of Table 4. The unconstrained multi-sample model provided a good fit to the data, indicating that the two-factor model provided adequate fit across both age groups. In the hierarchy of invariance constraints, neither constraining the factors loadings to equality across groups, χ^2 difference (10, N = 273) = 16.31, p >.05, nor constraining the correlation between factors and between residuals to equality across groups, χ^2 difference (12, N = 273) = 20.13, p > .05, resulted in significant reductions in model fit, compared to the fully unconstrained model. However, when all of the residuals were constrained to equality across groups, the model provided a significantly worse fit to the data than did the fully unconstrained model, χ^2 difference (22, N = 273) = 58.69, p < .001.

Sequential examination and release of the invariance constraints responsible for the most model misspecification indicated that the release of invariance constraints for four residuals (i.e., Bird Dragon, Size Ordering, Picture Imitation, Knock Tap) and one factor loading (i.e., Block Sorting) resulted in a model that fit the data as well as the fully unconstrained model, χ^2 difference (17, N = 273) = 20.45, p > .25. Therefore, whereas the same two-factor model provided an adequate fit to the structure of the data for both younger and older children, the degree to which the Block Sorting task indexed the IC factor and the degree to which scores on the Bird Dragon, Size Ordering, Picture Imitation, and Knock Tap tasks were accounted for by the model varied between younger and older children[2].

Given prior findings of unidimensional EF in preschool samples, a final multi-sample model comparison was conducted in which only the correlation between WM and IC was constrained to equality between younger and older children. The imposition of this constraint resulted in a significant reduction in model fit, χ^2 difference (1, N = 273) = 3.98, p < .05, and revealed that the correlation between WM and IC was significantly higher for younger preschool children (φ = .95) than it was for older preschool children (φ = .68).

Multi-sample models for girls versus boys—Descriptive statistics for task performance for the girls and boys in the sample are shown in Table 3. As shown in the table, there was only a single difference in observed EF task performance between girls and boys. Girls (M = 2.77, SD = 0.28) had higher scores on the Block Sorting task than did boys (M = 2.64, SD = 0.42), F(1, 266) = 9.88, p = .002. A summary of the multi-sample analyses comparing girls and boys is shown in the lower panel of Table 4. The unconstrained multi-sample model provided a good fit to the data, indicating that the two-factor model provided adequate fit for girls and boys. In the hierarchy of invariance constraints, neither constraining the factor loadings to equality across groups, χ^2 difference (10, N = 279) = 7.44, p > .50, nor constraining the correlations between factors and between residuals to equality across groups, χ^2 difference (12, N = 279) = 8.75, p > .50, resulted in a significant reduction in model fit, compared to the fully unconstrained model. When all of the residuals also were constrained to equality across groups, the fully constrained model did not provide a significantly worse fit than fully unconstrained model, χ^2 difference (22, N = 279) = 17.29, p > .50. Therefore, the same two-factor model provided adequate fit for girls and boys in the sample, and all parameters were equal in both groups.

[2]For the multi-sample models using the sandwich estimator to adjust standard errors to account for the clustering of children within schools, fewer residuals were found to differ between groups (Size Ordering and Bird Dragon only), and the difference in the factor loading for Block Sorting emerged during the step in which all 10 factor loadings were constrained simultaneously.

In light of evidence from observer ratings of effortful control of attention in general (see Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006, for review) and behavioral measures of IC in particular (Kochanska et al., 1996) that IC may be more developed among girls than boys during preschool, a final multi-sample model comparison was conducted to test for possible sex differences in the emergence of a separate IC factor. In this model, only the correlation between WM and IC was constrained to equality between girls and boys. The imposition of this constraint did not result in a significant decrease in model fit, χ^2 difference (1, N = 279) = .70, $p > .40$, indicating the same degree of overlap between WM and IC for girls and boys in this sample.

Discussion

Results of this study indicate that preschool children's EF is a multi-dimensional construct that includes distinct IC and WM components. The two-factor model in which tasks designed to measure WM and tasks designed to measure IC defined separate but correlated factors provided the best fit to the data, and this model provided a significantly better fit than a one-factor model for the sample as a whole. This finding represents the first evidence of divergence between component EFs among preschool children in a typically developing sample and indicates that the structure of EF in older preschool children is similar to that of at least the WM and IC components included in models of EF among adults and older children. Specifically, the uniformly significant and moderate-to-high loadings of EF tasks on separate factors closely parallel the WM and IC portion of findings reported by Miyake et al. (2000) among adults and by Lehto et al. (2003) among older children, as does the similar degree of overlap between the WM and IC factors. These results demonstrate the presence of separate WM and IC components among preschool children; thus, previous findings of unidimensionality of EF among preschool populations may have been the result of the methods or populations used in prior studies.

Several studies have examined the dimensionality of EF in typically developing preschool-age populations; however, none of these studies has produced evidence of multidimensionality of EF (e.g., Hughes & Ensor, 2007; Welsh et al., 2010; Willoughby et al., 2010; Wiebe, Espy, & Charak, 2008; Shing et al., 2010, Wiebe et al., 2011; but see Schoemaker et al., 2012). In contrast, results of this study support WM and IC as distinct dimensions of EF in typically developing preschool-age children. Differences between this and previous studies, including (a) improved construct specification and task selection, (b) confirmatory as opposed to exploratory factor analysis, (c) a larger number of tasks utilized to index WM and IC, and (d) purposefully focused age range of participants may explain why divergence between WM and IC has not typically been detected among preschool students.

Improved construct specification is among the most likely explanations for the unique finding of this study because it results in consistent application of the definitions of each component EF and makes possible the identification of latent variables directly comparable to those used with populations of older children and adults. To the extent that previous studies have used single tasks or composite scores derived from groups of tasks to form observed variables, task-specific variance cannot be separated from construct specific variance, thus preventing a direct examination of the boundary between the constructs of WM and IC. In this study, IC was represented by variance shared among six tasks, each of which required the suppression of a predisposed response and three of which also required the performance of a non-predisposed response. WM was indexed by variance shared across four tasks, each of which required either updating or manipulation of information held in memory. In an effort to be directly comparable to previous studies of EF in older populations, tasks used to index WM in this study fit a strict definition of requiring recall and either updating or manipulation of information held briefly in memory, as opposed to including forward span tasks, which require recall but neither updating nor manipulation. Because this study used the same definitions as studies of EF conducted with adult samples, the finding of the same pattern of divergence between WM and IC indicates the presence of two of the same component EFs among preschool children that, with one exception (Schoemaker et al., 2012), had previously been demonstrated only in samples of older children and adults.

Several of the tasks utilized in this study were used in previous studies that reported a single EF factor (e.g., Luria's Hand Game, Day-Night; Hughes & Ensor, 2007; Knock-Tap and Word Span Reversed; Welsh et al., 2010). To the extent that tasks used in this study differed from those used in previous examinations of the boundary between WM and IC among preschool children, these differences exist because tasks in this study were selected to fit the same component EF definitions used to examine the boundary between WM and IC among older children and adults. Therefore, to the extent that previous studies used tasks that did not fit the same definitions used with older children and adults, it is unlikely that distinctions between component EFs demonstrated among older populations would be replicated in a preschool sample. In fact, the only previous study to use CFA and apply the same definitions used with adults (Schoemaker et al., 2012) also demonstrated a boundary between WM and IC in a sample of preschool children diagnosed with ADHD.

Measurement and Structural Invariance of EFs Among Subgroups

The absence of any differences when comparing girls and boys indicates that WM and IC have the same meaning as EF components for both groups and that the relation between these components does not differ with regard to sex. When comparing younger and older children, each set of tasks indexed the same underlying skill in both groups. Although the two-factor model provided good fit in the overall sample, the degree of divergence between the WM and IC factors was significantly higher among older children than among younger children, indicating a greater degree of distinction between component EFs among older children. This pattern is consistent with the idea that EF components diverge as a function of age and adds to the body of evidence indicating rapid changes in EF that take place during the preschool years.

The finding that the distinctiveness of IC and WM increases across the preschool period may partially explain why previous studies of the dimensionality of EF in preschool children reported a unitary structure of EF. Although several of these earlier studies included mostly 3-year-old children, three of the five previous factor analytic studies that reported unitary EF among preschool students had samples of mostly 4-year-old children (i.e., Wiebe, Espy, &

Charak, 2008; Hughes & Ensor, 2007; Welsh et al., 2010). The average ages of children in these three studies ranged from 3.92 years (*SD* = 1.00) to 4.49 years (*SD* = .31), and the average age of children in the current study was 4.65 years (*SD* = .63). To the extent that studies include children at the lower end of the preschool age range, a distinction between IC and WM is less likely to emerge, particularly when less optimal measurement and analytic procedures are used.

One plausible explanation for the increasing distinctiveness of IC and WM across the preschool period is the result of increasing WM capacity. IC tasks require children to do something that is different than the prepotent response (e.g., withhold a response to a verbal command under some conditions; provide the opposite label for pictures). The ability to perform such responses requires children to hold the task rule in memory while responding to the task (i.e., attending to verbal commands or visual stimuli). Children with less WM capacity are less able to both attend to the task and recall the task rule than are children with more WM capacity. Therefore, the ability to respond correctly to the demands of an IC task may be equal to WM capacity until some threshold of WM capacity is achieved. Future studies should further explore the age and developmental boundaries for the emergence of distinct WM and IC capacities of young children, in addition to examining increasing WM capacity as a possible explanation for the initial divergence between WM and IC.

Subtypes of Inhibitory Control

Subtypes of IC were predicted based on the finding (Espy & Bull, 2005) that IC tasks requiring children to choose between multiple responses (i.e., conflict tasks) were significantly related to memory task performance, but IC tasks requiring only the suppression of a predisposed response (i.e., suppression tasks) were not related to memory performance, which the authors interpreted as an indication that IC tasks requiring the resolution of response conflict are more dependent on memory than IC tasks requiring only the inhibition of a predisposed response. There are several possible reasons for the absence of a similar distinction in this study. One possibility is that the response conflict tasks in this study only required children to choose between two stimulus-response options, whereas the tasks used by Espy and Bull appeared to place a larger demand on memory. The difference could also be due to the fact that this study used working memory tasks as opposed to forward span tasks. Alternatively, it is possible that other task demands distinguished the two groups of inhibitory tasks utilized by Espy and Bull, given that observed variables--as opposed to latent variables--were used and that the range of tasks differed by design on a number of characteristics.

Importance of Executive Functions During Preschool

An increasing body of evidence supports the connection between EF deficits and a range of behavioral problems, including aggressive behavior (e.g., Bierman, Torres, Domitrovich, Welsh, & Gest, 2009) and internalizing as well as externalizing symptoms (e.g., Rhoades et al., 2009). The importance of examining component EFs separately was underscored by Raaijmakers et al. (2008) who reported that children rated as high in aggressive behaviors differed from their peers on IC tasks but not on other EF tasks. As suggested by Blair, Zelazo, and Greenberg (2005), an improved understanding of early EFs will enable more direct examination of the differences between WM and IC as predictors of a range of outcomes, as well as more precise identification of the causes of poor behavioral regulation in general and externalizing symptoms in particular. Such examination can extend and clarify the meaning of findings that currently suggest WM and IC relate strongly, and in some cases differently, to academic and socio-emotional outcomes.

Limitations

Although a distinction between WM and IC dimensions was identified, the current study was not designed to address the additional question of whether or not a distinct SH dimension can be identified in a preschool population. Because SH was not measured, this study cannot address questions regarding the development of SH as an EF component among preschool children. It is possible that a different preferred model may have been identified if a SH factor had been included; however, five of the six previous factor analytic studies that reported the absence of a boundary between WM and IC also lacked a SH measure. As a result, the emergence of a boundary between WM and IC was unlikely to be the result of the exclusion of SH measures.

Scores on several of the IC tasks were characterized by moderate levels of negative skew, indicating that these commonly used tasks may be insufficiently difficult to capture the high end of IC task performance among 4- and 5-year-old children. These tasks might be revised to include a wider range of difficulty so as to remain sensitive to individual differences present at higher and lower levels of IC performance. In light of evidence from simulation studies, the moderate levels of skew observed in this study were not sufficient to result in model misspecification.

Another possible limitation was that the responses required for all WM tasks were verbal, whereas the responses required for all IC tasks were motor. Such systematic differences in task requirements could potentially lead to the appearance of separate factors; however, this explanation is unlikely for several reasons. First, the fact that studies reporting a single EF factor have used tasks with similar response differences (i.e., motor for IC and verbal for WM) makes it unlikely that this difference between tasks could account for the divergence between factors. Second, results of model comparisons that controlled for children's oral language skills supported the two-factor model, showing that the distinction between WM and IC factors exists independent of any differences in language skill required by each type of task. Finally, given that younger and older children completed exactly the same tasks with exactly the same instructions, if differences in response type were responsible for the observed divergence between factors in the overall sample, the degree of divergence would have been equal when comparing these groups. Instead, significantly more divergence between EF components was observed among older children than among younger children. This pattern could not be explained by response type differences that were identical for both groups.

Future Directions

Although divergence between WM and IC was observed among preschool children, this evidence cannot be used to specify the developmental timing of such divergence. Relatedly, despite the evidence of separate EF components among both older and younger children in this sample, the younger group was not large enough to permit a separate test of EF factor structure. Such a test would require a larger sample with age concentrated around the point at which separable components are expected to emerge. Future studies should seek to determine the developmental timing of the emergence of separate EF components by including a wider age range. Performance on EF tasks relates significantly to academic and socio-emotional outcomes in preschool children and has been linked to the development of psychopathology (Nigg & Casey, 2005), but the mechanism underlying these relations is not well understood. A clearer understanding of the dimensionality of EF would inform more specific hypotheses regarding the connection between EF components and a wide range of outcomes. Any test of component-specific relations using preschool EF as a predictor would require the measurement of separate component EFs during preschool. Otherwise, the examination of specific predictive relations could only begin later in development, when outcomes of interest may be less malleable and maladaptive behaviors more pervasive.

Acknowledgments

Portions of this work were supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD052120) and the Institute of Education Sciences, US Department of Education (R305B04074). The views expressed are those of the authors and have not been reviewed or approved by the granting agencies. Neither author is aware of any conflict of interest regarding the findings of this study.

References

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA preschool forms and profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families; 2000.

Alloway TP, Gathercole SE, Kirkwood H, Elliott J. The cognitive and behavioral characteristics of children with low working memory. Child Development. 2009; 80:606–621. [PubMed: 19467014]

Bierman KL, Torres MM, Domitrovich CE, Welsh JA, Gest SD. Behavioral and cognitive readiness for school: Cross-domain associations for children attending head start. Social Development. 2009; 18:305–323.

Blair C, Zelazo PD, Greenberg MT. The measurement of executive function in early childhood. Developmental Neuropsychology. 2005; 28:561–571. [PubMed: 16144427]

Brocki KC, Eninger L, Thorell LB, Bohlin G. Interrelations between executive function and symptoms of hyperactivity/impulsivity and inattention in preschoolers: A two year longitudinal study. Journal of Abnormal Child Psychology. 2010; 38:163–171. [PubMed: 19763816]

Bull R, Espy KA, Wiebe SA. Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental Neuropsychology. 2008; 33:205–228. [PubMed: 18473197]

Bull R, Scerif G. Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. Developmental Neuropsychology. 2001; 19:273–293. [PubMed: 11758669]

Carlson SM, Moses LJ. Individual differences in inhibitory control and children’s theory of mind. Child Development. 2001; 72:1032–1053. [PubMed: 11480933]

Corrigan R. The effects of task and practice on search for invisibly displaced objects. Developmental Review. 1981; 1:1–17.

Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods. 1996; 1:16–29.

Elliott, CD. Differential Abilities Scale. San Antonio, TX: Psychological Corporation; 1990.

Else-Quest NM, Hyde JS, Goldsmith HH, Van Hulle CA. Gender differences in temperament: A meta-analysis. Psychological Bulletin. 2006; 132:33–72. [PubMed: 16435957]

Espy KA. The Shape School: Assessing executive function in preschool children. Developmental Neuropsychology. 1997; 13:495–499.

Espy KA, Bull R. Inhibitory processes in young children and individual variation in short-term memory. Developmental Neuropsychology. 2005; 28:669–688. [PubMed: 16144432]

Frye D, Zelazo PD, Palfai T. Theory of mind and rule-based reasoning. Cognitive Development. 1995; 10:483–527.

Gathercole SE, Pickering SJ. Assessment of working memory in six- and seven-year-old children. Journal of Educational Psychology. 2000; 92:377–390.

Gerstadt CL, Hong YJ, Diamond A. The relationship between cognition and action: Performance of children 3 1/2-7 years old on a stroop-like day-night test. Cognition. 1994; 53:129–153. [PubMed: 7805351]

Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999:1–55.

Hughes C. Control of action and thought: Normal development and dysfunction in autism: A research note. Journal of Child Psychology and Psychiatry. 1996; 37:229–236. [PubMed: 8682904]

Hughes C, Ensor R. Executive function and theory of mind: Predictive relations from ages 2 to 4. Developmental Psychology. 2007; 43:1447–1459. [PubMed: 18020823]

Hughes C, White A, Sharpen J, Dunn J. Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. Journal of Child Psychology and Psychiatry. 2000; 41:169–179. [PubMed: 10750543]

Johnson, MH. Developmental cognitive neuroscience. 2nd ed.. Oxford, England: Blackwell Publishers; 2005.

Jones LB, Rothbart MK, Posner MI. Development of executive attention in preschool children. Developmental Science. 2003; 6:498–504.

Kerns KA, Rondeau LA. Development of a continuous performance test for preschool children. Journal of Attention Disorders. 1998; 2:229–238.

Kline, RB. Principles and practice of structural equation modeling. 3rd ed.. New York: The Guilford Press, NY; 2011.

Kochanska G, Murray K, Jacques TY, Koenig AL, Vandegeest KA. Inhibitory control in young children and its role in emerging internalization. Child Development. 1996; 67:490–507. [PubMed: 8625724]

Kooijmans R, Scheres A, Oosterlaan J. Response inhibition and measures of psychopathology: A dimensional analysis. Child Neuropsychology. 2000; 6:175–184. [PubMed: 11402395]

Korkman, M.; Kirk, U.; Kemp, S. NEPSY: A developmental neuropsychological assessment. San Antonio, TX: Psychological Corporation; 1998.

Lehto JE, Juujärvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. British Journal of Developmental Psychology. 2003; 21:59–80.

McInerney RJ, Hrabok M, Kerns KA. The children's size-ordering task: A new measure of nonverbal working memory. Journal of Clinical and Experimental Neuropsychology. 2005; 27:735–745. [PubMed: 16019649]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology. 2000; 41:49–100. [PubMed: 10945922]

Muthén, LK.; Muthén, BO. Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén; 1998–2011.

Nigg JT, Casey BJ. An integrative theory of attention-deficit/hyperactivity disorder based on the cognitive and affective neurosciences. Development and Psychopathology. 2005; 17:785–806. [PubMed: 16262992]

Nigg JT, Quamma JP, Greenberg MT, Kusche CA. A two-year longitudinal study of neuropsychological and cognitive performance in relation to behavioral problems and competencies in elementary school children. Journal of Abnormal Child Psychology. 1999; 27:51–63. [PubMed: 10197406]

Pelphrey K, Reznick S, Goldman B, Sasson N, Morrow J, Donahoe A, et al. Development of visuospatial short-term memory in the second half of the 1st year. Developmental Psychology. 2004; 40:836–851. [PubMed: 15355170]

Pickering S, Gathercole SE. Working Memory Test Battery for Children (*WMTB-C*). Psychological Corporation. 2001

Raaijmakers MAJ, Smidts DP, Sergeant JA, Maassen GH, Posthumus JA, van Engeland H, Matthys W. Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. Journal of Abnormal Child Psychology. 2008; 36:1097–1107. [PubMed: 18437548]

Reed MA, Pien DL, Rothbart MK. Inhibitory self-control in preschool children. Merrill-Palmer Quarterly. 1984:131–147.

Reznick J, Morrow J, Goldman B, Snyder J. The onset of working memory in infants. Infancy. 2004; 6:145–154.

Rhoades BL, Greenberg MT, Domitrovich CE. The contribution of inhibitory control to preschoolers' social-emotional competence. Journal of Applied Developmental Psychology. 2009; 30:310–320.

Riggs NR, Blair CB, Greenberg MT. Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. Child Neuropsychology. 2003; 9:267–276. [PubMed: 14972705]

Schoemaker K, Bunte T, Wiebe SA, Espy KA, Deković M, Matthys W. Executive function deficits in preschool children with ADHD and DBD. Journal of Child Psychology and Psychiatry. 2012; 53:111–119. [PubMed: 22022931]

Seigneuric A, Ehrlich M-F. Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. Reading and Writing. 2005; 18:617–656.

Shing YL, Lindenberger U, Diamond A, Li S-C, Davidson MC. Memory maintenance and inhibitory control differentiate from early childhood to adolescence. Developmental Neuropsychology. 2010; 35(6):679–697. [PubMed: 21038160]

Wechsler, D. Wechsler Intelligence Scale for Children—4th Edition (WISC-IV®). San Antonio, TX: Pearson; 2003.

Welsh JA, Nix RL, Blair C, Bierman KL, Nelson KE. The development of cognitive skills and gains in academic school readiness for children from low-income families. Journal of Educational Psychology. 2010; 102:43–53. [PubMed: 20411025]

Wiebe SA, Espy KA, Charak D. Using confirmatory factor analysis to understand executive control in preschool children: I. latent structure. Developmental Psychology. 2008; 44:575–587. [PubMed: 18331145]

Wiebe SA, Sheffield T, Nelson JM, Clark CAC, Chevalier N, Espy KA. The structure of executive function in 3-year-olds. Journal of Experimental Child Psychology. 2011; 108:436–452. [PubMed: 20884004]

Willoughby MT, Blair CB, Wirth RJ, Greenberg M. The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. Psychological Assessment. 2010; 22:306–317. [PubMed: 20528058]

Yuan, K-H.; Bentler, PM. Mean and covariance structure analysis with missing data. In: Gupta, A.; Girko, V., editors. Multidimensional statistical analysis and theory of random matrices: Proceedings of Sixth Eugene Lukacs Symposium. Utrecht: VSP; 1996. p. 307-326.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Descriptive Statistics and Intercorrelations for Executive Function Tasks

Task	*M* (SD)	Skew	Kurtosis	1	2	3	4	5	6	7	8	9
1. Object Span	1.19 (1.11)	.70	−.09	--								
2. Size Ordering	2.42 (1.73)	.73	.09	.05	--							
3. Word Span Reversed	2.12 (1.97)	.32	−1.08	.20**	.38***	--						
4. Listening Span	6.61 (4.08)	−.19	−1.29	.27***	.27***	.24***	--					
5. Knock-Tap	2.57 (.55)	−1.72	−.07	.12*	.21***	.32***	.20**	--				
6. Block Sort	2.71 (.36)	−1.95	.44	.14*	.16*	.25***	.07	.18**	--			
7. Day Night	2.55 (.52)	−1.47	−.81	.15*	.32***	.30***	.30***	.33***	.16*	--		
8. Picture Imitation	2.51 (.86)	−1.94	.84	.11	.28***	.31***	.29***	.35***	.20**	.35***	--	
9. Bird Dragon	2.70 (.77)	−2.73	3.78	.08	.22***	.23***	.29***	.39***	.13*	.39***	.61***	--
10. Luria's Hand Game	2.15 (.88)	−1.06	−.81	.20**	.19**	.29***	.16**	.46***	.15*	.25***	.30***	.34***

Note. *N*s range from 264 to 274.

* $p < .05$;

** $p < .01$;

*** $p < .001$.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Indices of Absolute and Relative Model Fit for Models of Children's Performance on Executive Function Tasks

Models	Y-Bχ^2	*df*	CFI	TLI	RMSEA	AIC	ΔY-Bχ^2
Model comparison using raw EF scores							
One-Factor Model							
Model 1: WM + IC	71.67^{***}	34	.94	.93	.06	7571	---
Two-Factor Models							
Model 2: WM, IC	55.60^{**}	33	.97	.96	.05	7555	10.55^{**}
Model 3: WM + IC_C; IC_S	72.10^{***}	33	.94	.92	.07	7572	.19
Model 4: WM + IC_S; IC_C	69.75^{***}	33	.95	.93	.06	7570	1.92
Three-Factor Model							
Model 5: IC_c; IC_s; WM	55.88^{**}	31	.96	.95	.05	7558	.22
Model comparison using oral language-adjusted EF scores							
One-Factor Model							
Model 1: WM + IC	54.46^{*}	34	.88	.84	.05	7518	---
Two-Factor Models							
Model 2: WM, IC	44.45^{ns}	33	.93	.91	.04	7509	8.83^{**}
Model 3: WM + IC_C; IC_S	55.50^{**}	33	.87	.92	.05	7520	.04
Model 4: WM + IC_S; IC_C	53.55^{*}	33	.88	.93	.05	7518	.97
Three-Factor Model							
Model 5: IC_c; IC_s; WM	43.31^{ns}	31	.93	.89	.04	7511	1.42

Note. N = 284. χ^2-difference tests for two-factor models involve comparisons to Model 1, and χ^2-difference test for the three-factor model involved comparison to Model 2. WM = Working Memory, IC_S = Inhibitory Control, suppression subtype, IC_C = Inhibitory Control, response conflict subtype. Y-B χ^2 = Yuan-Bentler Chi-Square; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation; AIC = Akaike's Information Criterion.

$^{ns}p > .05$;

$^{*}p < .05$;

$^{**}p < .01$;

$^{***}p < .001$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Standardized Model Parameters for Overall Sample and Descriptive Statistics and Standardized Model Parameters for Younger and Older Children in Sample

	Overall Sample			Younger Children				Older Children				
	Factor Loadings				Factor Loadings				Factor Loadings			Group Mean comparison
Tasks	**WM**	**IC**	**R^2**	**Mean (*SD*)**	**WM**	**IC**	**R^2**	**Mean (*SD*)**	**WM**	**IC**	**R^2**	***p*-value**
Working Memory												
Word Span Reversed	.60		.36	1.72 (1.93)	.49		.24**	2.65 (1.92)	.70		.49	.001
Listening Span	.53		.28	6.15 (4.13)	.39		.16*	7.10 (4.01)	.57		.33	.06
Size Ordering	.55		.30	2.10 (1.41)	.49		.24*	2.82 (1.98)	.57		.32	.001
Object Span	.34		.11**	1.13 (1.03)	.22*		.05^{ns}	1.21 (1.17)	.44		.19*	.56
Inhibitory Control												
Bird & Dragon		.68	.47	2.56 (0.91)		.66	.44	2.86 (0.51)		.74	.54	.001
Luria's Hand Game		.60	.36	2.09 (0.95)		.64	.41	2.24 (0.79)		.52	.27	.17
Block Sorting		.57	.33**	2.69 (0.38)		.45*	.20^{ns}	2.70 (0.35)		.72**	.52**	.34
Day-Night Stroop		.71	.50	2.50 (0.57)		.67	.45	2.60 (0.47)		.76	.57	.12
Knock-Tap		.74	.55	2.51 (0.64)		.71	.50	2.61 (0.45)		.81	.66	.13
Picture Imitation		.64	.41	2.38 (1.00)		.61	.37	2.66 (0.66)		.64	.42	.01

Note. WM = working memory; IC = Inhibitory Control; R^2 = percentage of variance in task accounted for by factor. All model parameters were significant at $p \leq .001$ unless otherwise noted.

$^{ns}p > .05$;

$^{*}p < .05$;

$^{**}p < .01$.

Table 4

Indices of Absolute and Relative Model Fit for Multi-Sample Tests of Measurement and Structural Invariance between Older and Younger Preschool Children (upper panel) and Girls and Boys (lower panel)

Constraints	Y-Bχ^2	*df*	CFI	TLI	RMSEA	AIC	ΔY-Bχ^2
Multi-sample Models for Younger and Older Children							
None	104.90**	74	.96	.95	.06	7205	---
Factor Loadings	122.30***	84	.95	.94	.06	7210	16.31^{ns}
Factor Loadings + Covariances	126.81**	86	.94	.94	.06	7215	3.69^{ns}
Factor Loadings + Covariances + Residuals	176.00***	96	.89	.89	.08	7259	39.45***
Model with 5 Constraints Released	124.27*	91	.95	.95	.05	7199	11.14^{ns}
Multi-sample Models for Girls and Boys							
None	116.67**	74	.94	.93	.06	7457	---
Factor Loadings	120.07**	84	.95	.95	.06	7449	7.44^{ns}
Covariances + Factor Loadings	120.38**	86	.95	.95	.05	7447	1.32^{ns}
Covariances + Factor Loadings + Residuals	126.76*	96	.96	.96	.05	7441	8.56^{ns}

Note. *N* = 273 in multi-sample models for age, and *N* = 279 in multi-sample models for boys and girls. Y-Bχ2 = Yuan-Bentler χ^2; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Squared Error of Approximation; AIC = Akaike's Information Criterion. Covariances refers to covariance between Working Memory and Inhibitory Control factors and to covariance between the residuals of the Bird Dragon and the Picture Imitation tasks. Residuals refers to residual variances for individual tasks. ΔY-B χ^2 values reflect comparison to prior model.

$^{ns}p > .05$;

$^{*}p < .05$;

$^{**}p < .01$;

$^{***}p < .001$.