Journal of Experimental Child Psychology 108 (2011) 436–452

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The structure of executive function in 3-year-olds

Sandra A. Wiebe [a,*], Tiffany Sheffield [b], Jennifer Mize Nelson [b], Caron A.C. Clark [b], Nicolas Chevalier [b], Kimberly Andrews Espy [b]

[a] *Department of Psychology, University of Alberta, Edmonton, Alberta, T6G 2E9 Canada*
[b] *Department of Psychology and Office of Research, University of Nebraska–Lincoln, Lincoln, NE 68588, USA*

ARTICLE INFO

Article history:
Available online 29 September 2010

Keywords:
Executive function
Working memory
Inhibitory control
Preschool
Socioeconomic status
Confirmatory factor analysis
Measurement invariance

ABSTRACT

Although the structure of executive function (EF) during adulthood is characterized by both unity and diversity, recent evidence suggests that preschool EF may be best described by a single factor. The latent structure of EF was examined in 228 3-year-olds using confirmatory factor analysis. Children completed a battery of executive tasks that differed in format and response requirements and in putative working memory and inhibitory control demands. Tasks appeared to be age appropriate, with adequate sensitivity across the range of performance and without floor or ceiling effects. Tests of the relative fit of several alternative models supported a single latent EF construct. Measurement invariance testing revealed less proficient EF in children at higher sociodemographic risk relative to those at lower risk and no differences between boys and girls. At 3 years of age, when EF skills are emerging, EF appears to be a unitary, more domain-general process.

Introduction

The regulation of goal-directed behavior is supported by a number of mental processes, including monitoring and updating the contents of working memory, dampening or overriding prepotent responses, and flexibly shifting behavior depending on contextual demands (Miyake et al., 2000). These abilities are collectively referred to by several names in the literature, including (but not limited to) executive function, executive control, and cognitive control. The term *executive control* has a number of advantages in that it captures the common supervisory or self-regulatory nature of these processes ("executive") and the idea that this regulation is achieved by modulation of subordinate processes

* Corresponding author.
E-mail address: sandra.wiebe@ualberta.ca (S.A. Wiebe).

0022-0965/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2010.08.008

(“control”). However, because this article forms part of a special issue, we use the term *executive function* (EF) for consistency. The neural circuitry subserving EF is slow to develop, with prefrontal regions fully maturing only during early adulthood (Lenroot & Giedd, 2006; Paus et al., 1999; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999). For a long time, EF competence during early childhood was assumed to be negligible and thus of little or no relevance to understanding preschool behavior (Chelune & Baer, 1986) despite evidence of substantial development and reorganization in prefrontal systems during this period (Huttenlocher, 1990; Thatcher, 1992). It has since been recognized that simple behaviors emerging as early as infancy, such as the regulation of eye movements (Johnson, 1995) and searching for hidden objects (Diamond, 1990), likely represent the developmental antecedents of more complex executive skills.

Recent efforts to develop new paradigms to assess EF in preschool children have enabled substantial progress in understanding milestones in EF growth during this important period (for reviews, see Carlson, 2005; Espy, Kaufmann, McDiarmid, & Glisky, 1999; Garon, Bryson, & Smith, 2008). Cross-sectional studies of EF at the manifest level, using tasks that assess the putative EF components of working memory and inhibition, have suggested that 3 years of age may be pivotal for EF development. Rapid gains in performance on inhibition and delay of gratification tasks are observed between 3 and 5 years of age (Carlson & Moses, 2001; Diamond & Taylor, 1996; Kochanska, Murray, & Harlan, 2000).

Manifest performance on a given executive task, however, does not directly reflect the underlying construct of EF during the preschool years, as is often assumed. The relation between the latent EF construct and observed performance on different preschool tasks has not yet been sufficiently investigated. Progress on this front is critically important to better explicate developmental relations between EF and academic outcomes (Assel, Landry, Swank, Smith, & Steelman, 2003; Bull, Espy, & Wiebe, 2008), clinical symptomatology (Brocki, Nyberg, Thorell, & Bohlin, 2007; Raaijmakers et al., 2008), and social and biological risk (Noble, McCandliss, & Farah, 2007; Noble, Norman, & Farah, 2005; Taylor, Minich, Klein, & Hack, 2004). Furthermore, a more fine-grained understanding of EF organization will enable a more precise linkage of cognitive processes and biological correlates, for example, candidate genes or neurophysiology (e.g., Diamond, Briand, Fossella, & Gehlbach, 2004; Fassbender et al., 2004; Garavan, Ross, Murphy, Roche, & Stein, 2002; McNab et al., 2008; Wiebe et al., 2009). The preschool years are a critical transition period as rapid changes in language ability, symbolic thought, and self-understanding enable better regulated and more goal-directed behavior (Carlson, 2005; Espy et al., 1999). An accurate conceptualization of EF structure at this key stage of development is critical to the design of effective and efficient preventive interventions to benefit children with early difficulties (Diamond, Barnett, Thomas, & Munro, 2007).

Despite these clear advantages of a more precise conceptualization of early EF, the characterization of its structure in very young children presents several challenges. EF is a tertiary process by definition; that is, it regulates other modular abilities such as language, visual–spatial, motor, and memory skills. Thus, individual differences in performance on any single executive task result not only from variation in EF proficiency but also from variation in these other nonexecutive abilities that are necessary to perform the task. This issue is commonly referred to as the “task impurity” problem (Miyake et al., 2000). Because early childhood is characterized by substantial variation in acquisition of these nonexecutive skills, the potential confounding influence of variability in nonexecutive skills on EF measurement is of particular concern during the preschool years.

To better estimate EF separately from the influence of nonexecutive skills, a number of recent studies have used confirmatory factor analysis (CFA) (e.g., Friedman & Miyake, 2004; Miyake et al., 2000; Wiebe, Espy, & Charak, 2008). In this approach, a battery of tasks that share a requirement for EF but differ in other stimulus and response demands are administered to the same sample, thereby making it possible to parse common EF variance from that attributable to other sources. Task performance scores are then modeled as indicators of an underlying latent variable (or variables), with the resulting latent construct representing a “purified” EF measure. Across multiple studies using CFA in adults, the best-fitting model includes three factors typically labeled inhibition, updating, and shifting. These constructs are distinct but strongly correlated, with correlations typically ranging from .40 to .74 (Friedman et al., 2007; Friedman et al., 2008; Miyake et al., 2000). Based on these findings, the diversity of EF, at least from school age onward, is now broadly acknowledged (but see McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010, for an exception). The CFA approach to studying EF has proven to be tremendously productive during adulthood (Friedman & Miyake, 2004; Miyake et al., 2000), adolescence (Friedman et al., 2007; Friedman et al., 2008), and childhood (Huizinga, Dolan, & van der Molen, 2006; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003) and has resulted in novel insights. For example, Friedman and colleagues (2008) combined latent variable measurement of EF with a behavior genetics approach in a sample of adolescent twins to parse the variability in EF and its subcomponents into that attributable to genetic variation, shared environmental effects, and unshared environmental effects. Surprisingly, variability in latent EF was almost entirely attributable to genetic variation. In the same sample, attention problems during childhood were related to latent EF during adolescence (Friedman et al., 2007).

Although the EF structure that reflects both unity and diversity appears to be applicable from middle childhood onward, the picture during early childhood differs. Wiebe and colleagues (2008) found that, in a sample of children ranging from 3 to 6 years of age, one factor best fit the observed data model, and modeling separate working memory and inhibition factors did not improve model fit. The unitary factor was invariant across boys and girls as well as between preschoolers whose mothers had achieved a high school education and children whose mothers had completed some college or more. Girls and older children had higher mean latent EF, whereas there were no latent mean differences between groups of children whose mothers differed in educational background. In addition, there was some support for invariance across age, although these analyses were interpreted cautiously because there was a higher prevalence of missing data in younger children. Recently, Hughes, Ensor, Wilson, and Graham (2010) found that a unitary model of EF evidenced a good fit to preschoolers' EF task performance in a longitudinal study where children were assessed at 4 and 6 years of age using a task battery that included measures of inhibitory control, working memory, and planning. Similar to Wiebe and colleagues (2008), Hughes and colleagues (2010) demonstrated adequate measurement invariance for sex and age, with better EF in older children. In contrast to the findings by Wiebe and colleagues, there were no sex differences in latent EF. Of note, this unitary model of preschool EF has proven relevance beyond the laboratory, where latent unitary EF is related to problem behaviors (Espy, Sheffield, Wiebe, Clark, & Moehr, in press), and emerging math skills (Bull, Espy, Wiebe, Sheffield, & Nelson, in press). This unitary model during early childhood can be reconciled with diversity of executive processes later in development if the structure of these mental abilities changes with development. There is evidence for such structure differentiation with advancing development for other process domains. For example, using factor analytic techniques, Tucker-Drob (2009) recently showed that relationships between different components of general intellectual ability became more clearly differentiated with age. Garon and colleagues (2008) proposed that the components of EF emerge in sequence across the preschool years, with working memory coming online first and followed by inhibition, which together enable the development of shifting, although this sequence has not been tested empirically. Hughes (2002) argued that understanding EF structure early in development, when it is emerging, has the potential to provide key insights into fundamental executive processes because tasks must be simpler with fewer extraneous nonexecutive demands that might cloud interpretation. Clearly, adequate measurement of preschool EF is a necessary step in determining whether the structure of EF changes across early and middle childhood.

When choosing appropriate tasks for young preschoolers, a number of other unique issues come into play. Preschool samples are characterized by variability in attention span, linguistic competence, and general background knowledge. Consequently, preschool tasks must be designed to minimize the complexity of verbal or manual responses and rules to be learned. Tasks must draw on basic concepts that children from all backgrounds might be expected to have mastered by this age. Tasks have typically drawn on children's knowledge of basic shapes, colors, animals, and common household objects (e.g., Espy, 1997; Hughes, Dunn, & White, 1998; Prevor & Diamond, 2005; Zelazo, Frye, & Rapus, 1996) and have used task frameworks that are familiar to children from their everyday experience, for instance, searching for hidden objects (e.g., Espy et al., 1999), waiting for a delayed reward (Kochanska et al., 2000), or imitating another's actions (Diamond & Taylor, 1996). For the current study, tasks that met a number of criteria were selected. Tasks needed to have been used previously with young preschoolers, with demonstrated reliability and validity, and were required to vary in nonexecutive task demands. Because this study represents the initial time point of a longitudinal investigation of EF, all tasks were pilot-tested with children representing the entire age range of interest before final selection to ensure applicability to the study population.

The core feature of a working memory task is the requirement to hold information in mind and use it to guide behavior. Several tasks that used an object search paradigm were selected. In the Nine Boxes task (Diamond, Prevor, Callender, & Druin, 1997), children needed to update their mental representations of remaining hidden toys from trial to trial so as to search with maximum efficiency, and in the Delayed Alternation task (Espy et al., 1999), children needed to update and maintain their representation of the treat's location from trial to trial across a filled delay. (More comprehensive descriptions of the tasks are given in the Method section.) Another tradition in the measurement of working memory is span tasks. Simple forward span tasks typically are considered to measure short-term memory rather than working memory. In preschoolers, backward span measures yield floor-level performance or missing data because of difficulty with task instructions or the concept of "backward" (e.g., Bull et al., 2008). A potential solution is the integration of forward span with a requirement to manipulate information (e.g., in complex span tasks). In the Noisy Book task (Hughes et al., 1998), children needed to remember a sequence of animal names and respond by pressing different unlabeled buttons that corresponded to the animals. All of these tasks had a multitrial structure, thereby requiring children to manage proactive interference.

To assess inhibitory control, the essential requirement is that children suppress a prepotent response. This competing response could have its origin in children's own behavioral repertoire (e.g., children's natural preference to reach for a reward immediately rather than delaying their response), as in the Snack Delay task (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). Alternatively, the competing response could be primed by the presence of distractors in the stimulus array, as in the Big–Little Stroop (Kochanska et al., 2000), where children must name the small pictures rather than the larger and more salient picture. In the Shape School task (Espy, 1997), the prepotent color-naming response is well practiced during an earlier phase of the task, making it more difficult to suppress this response on a subset of trials during the Inhibit phase of the task. Finally, in the Go/No-Go task (Simpson & Riggs, 2006), the task is structured to promote a bias to respond by requiring children to respond on the majority of trials.

For the third core component of EF, shifting or cognitive flexibility, the essential requirement is that children shift between tasks or mental sets. Although several shifting tasks were included in the study battery, data for one task were not yet coded, and because the Inhibit phase of the Shape School task was used to index inhibition, including the switch phase would have been problematic. Furthermore, 3-year-olds typically perseverate, showing a substantial inability to switch from one task set to another (e.g., Chevalier & Blaye, 2008; Hanania, 2010; Zelazo et al., 1996).

Because a longer term goal of this program of research is to elucidate individual pathways to key outcomes, the role of child biological and experiential factors and their potential impact on the fundamental structure of EF and its development require further explication. Two factors that may be of particular relevance are sex and family socioeconomic background or sociodemographic risk. Previous studies in preschoolers have provided some evidence that girls have a modest advantage in latent EF relative to boys (standardized difference = .89 [Wiebe et al., 2008]) and that girls outperform boys on some executive tasks, particularly for effortful control tasks involving a delay of gratification (Carlson & Moses, 2001; Kochanska et al., 2000; Matthews, Ponitz, & Morrison, 2009; Olson, Sameroff, Kerr, Lopez, & Wellman, 2005). In contrast, other studies have failed to find sex differences in EF task performance (Deák, Rey, & Pick, 2004; Hughes & Ensor, 2005). Higher sociodemographic risk may be associated with poorer performance on EF measures in young children (Allhusen et al. 2005; Mezzacappa, 2004; Noble et al., 2005; Noble et al., 2007). However, studies examining differences at the manifest level assume structural invariance (i.e., that a task measures the same construct across groups being compared). Only two studies to date have examined associations between these influences and preschool EF at the latent level, with these studies reporting conflicting results. Contrary to the findings of Wiebe and colleagues (2008) discussed above, Hughes and colleagues (2010) reported no differences in the latent structure or level of executive task performance in girls and boys but found that sociodemographic risk, indexed by family income, was related to the level of latent EF (B = .62) but not to its change with age. Given these inconsistent findings and the paucity of research related to individual differences related to sex or sociodemographic risk, these questions were revisited in the current study. The latent variable approach is particularly useful to address this question because group differences in the latent EF construct can be separated from differences in extraneous nonexecutive factors (e.g., color knowledge, verbal skill) that can affect task performance at the manifest level and that are known to vary by sociodemographic risk or sex (e.g., Farah et al., 2008; Magnuson, Sexton, Davis-Kean, & Huston, 2009; Wallentin, 2009).

The goal of the current study was to evaluate the fit of unitary and fractionated models of EF in young preschoolers at the beginning of this important and dynamic developmental period, namely, 3 years of age. CFA studies to date have been weakest in their ability to adequately capture EF in young preschoolers, largely because of a scarcity of tasks suitable for this young age. Taking advantage of the first phase of a longitudinal study of preschool EF development, we focused on determining EF structure in children tested within a narrow time window, namely, within 3 weeks of their third birthday. Adequately characterizing EF before this period of change provides a baseline for future work at a time point when EF structure is most likely to differ from its structure during adulthood based on previous findings supporting unitary preschool EF (Wiebe et al., 2008). Executive tasks were selected to yield multiple indicators of working memory and inhibitory control, which also differed in other respects (e.g., task format, use of rewards, verbal demands, spatial information) to enable the systematic testing of alternative models and to determine the role of "secondary" task factors. Descriptive statistics and distributional properties of the tasks were examined first to evaluate psychometric properties and adequacy of measurement in young preschoolers. Then the relative fits of a series of CFA models were evaluated to determine which model best represented the observed data patterns. Finally, invariance testing was used to investigate systematic differences in the latent EF related to sex and sociodemographic risk.

Method

Participants

The sample consisted of 228 3-year-olds (115 girls and 113 boys, mean age = 3.01 years, *SD* = 12.8 days). Participants were recruited through flyers distributed at local preschools, doctor's offices, and the local health department, as well as by word of mouth, from two midwestern U.S. study sites: a small city and a rural tricounty area. Before enrollment in the study, parents completed a telephone screening, and children with diagnosed developmental or language delays or behavioral disorders, whose primary language spoken in the home was not English, or whose families planned to move out of the area within the study timeline were not considered further for recruitment. To maintain consistency within the larger scale project, children diagnosed with developmental or language disorders later during longitudinal follow-up also were excluded from the current analyses.

The sample consisted of 173 Caucasian, 13 African American, 17 Hispanic, and 25 multiracial children. Because of the interest in understanding EF structure in different subgroups of children, sampling was stratified by sex and sociodemographic risk. A wide range of socioeconomic and educational backgrounds were represented in the sample. Children considered to be at sociodemographic risk, as defined by eligibility for public medical assistance (and labeled "at risk"), were oversampled so as to obtain roughly comparable numbers of participants in the low-risk (58%) and at-risk (42%) groups. On average, mothers had completed 14.7 years of education (range = 9–22, $M_{\text{low risk}} = 15.3$, $M_{\text{at risk}} = 13.8$).

Procedures

At a home visit, the parent provided written informed consent for the child's study participation. Approximately 1 week later, the child and parent (usually the mother, so hereafter referred to as the mother) came to the university laboratory, where a trained research technician administered the battery of executive tasks to the child in a child-friendly laboratory playroom. As is traditional in individual difference designs, tasks were administered in a fixed order (available from the first author), so that possible fatigue or transfer effects would be constant across participants. Prior to study initiation, pilot testing was conducted to determine an order of administration that maintained young preschoolers' attention and interest to maximize data completeness. The mother remained in the same room as her child for all tasks except Snack Delay so as to minimize any possible separation anxiety. During the child testing, the mother completed a background interview and a packet of rating forms related to the child's behavior and home environment.

A battery of executive tasks, varying in format and response demands, was administered to all children. These tasks were traditionally considered to reflect theoretical dimensions of working memory and inhibitory control. Several tasks were administered on the computer using E-Prime 1.1 (Psychology Software Tools, Pittsburgh, PA, USA), SuperLab (Cedrus, San Jose, CA, USA), or Perl v5.8.8 (ActiveState Software, Vancouver, BC, Canada). For the Go/No-Go task, children responded via button press; the remaining tasks were scored by the examiner during the session or later offline by a behavioral coder from video. For any tasks that required scoring, 20% of sessions were selected randomly and scored offline by trained undergraduates, who were naive to study hypotheses and sociodemographic risk status, to assess interrater reliability (reported below). Because most tasks yielded multiple dependent measures, the specific variables selected for inclusion in the CFA were determined on the basis of face validity for the intended construct and their distributional properties (i.e., ample variance without floor effects). All dependent measures considered for each task are listed with their distributional properties in Table 1, and the variable chosen for analysis from each task is described below. Although measures of cognitive flexibility were also included in the administered battery, there were obvious floor effects in this young age group that precluded inclusion in the current report.

Three tasks—Nine Boxes, Nebraska Barnyard, and Delayed Alternation—were considered a priori to place greater demands on working memory. In the Nine Boxes task (adapted from Diamond et al., 1997), children were required to search for figurines hidden in nine boxes with lids of varying shapes

Table 1
Descriptive statistics for all EF dependent measures.

Task and dependent measure	*n*	*M*	*SD*	Range	Skewness	Kurtosis
Nine Boxes						
Longest correct run	228	4.57	1.58	2–9	0.72	0.12
Efficiency score (correct reaches/total reaches)	228	.58	.13	.33–1	0.96	0.95
Nebraska Barnyard						
Summary score	220	1.16	0.61	0–3.3	0.59	0.52
One-item span	220	1.43	0.70	0–4	0.74	0.56
Two-item span	220	1.10	0.45	0–3	0.74	2.60
Delayed Alternation						
Longest correct run–longest incorrect run	227	−0.38	3.62	−12.26–11.54	−0.32	2.15
Number of correct responses	227	7.68	2.92	0–16	−0.05	0.11
Number of errors	227	8.22	2.93	0–16	0.04	0.13
Longest correct run	227	3.15	2.20	0–16	2.88	12.83
Big–Little Stroop						
Proportion correct (conflict trials)	222	.29	.29	0–1	0.98	−0.19
Proportion correct (controlling for matching trials accuracy)	222	−.21	.30	−.60–.66	0.60	−0.53
Response time for correct conflict trials (s)	175	3.37	4.05	0.03–34.89	4.57	26.85
Go/No-Go						
D'	221	0.58	0.52	0–2.39	1.36	1.57
Proportion correct (No-Go condition)	221	.48	.34	0–1	0.15	−1.38
Shape School (Inhibit condition)						
Proportion correct (suppression stimuli)	195	.36	.42	0–1	0.62	−1.36
Snack Delay						
Summary score	214	35.32	24.18	0–137	0.64	0.95
Number of epochs without rule violations	214	1.23	4.05	0–41	6.20	49.50
Ate snack during delay (yes = 1, no = 0)	214	.29	.46	0–1	0.91	−1.19
Latency to eat snack (s)	214	186.66	84.35	5–240	−1.20	−0.23

Note. For each task, the dependent measure selected for CFA analysis is italicized.

and colors. Children were allowed to open only one box per trial, with boxes being scrambled behind a screen during the 15-s delay between trials. A maximum of 20 trials were administered until children found all figurines or made five consecutive errors. The longest run of consecutive correct responses was the dependent variable selected for this task. All children had complete data on this task. Interrater reliability for the scoring of this task was 100%.

The Nebraska Barnyard task (adapted from the Noisy Book task of Hughes et al., 1998) is a computerized complex span task requiring children to remember a sequence of animal names and press corresponding buttons on a touch screen in the correct order. During an initial training phase, children were introduced to a set of nine colored pictures of barnyard animals arranged in a 3 × 3 grid of colored "boxes" on the computer screen. As children pressed each animal box, the computer produced a corresponding animal sound. Children who could not name the animals for the majority of these trials were given a score of 0 ($n = 7$). Thereafter, the animal pictures were replaced with blank boxes, where the box colors were associated with animal identity when possible (e.g.., the "frog" button was green, the "cow" button was brown), although in some cases the association was weaker (e.g., the "cat" button was orange, the "horse" button was black). Children completed a set of nine trials during which the examiner named each animal individually and children were required to press the box corresponding to that animal. Finally, trials with sequences of animals were administered, beginning with sequences of two animals and increasing progressively until children's performance deteriorated. Up to three trials were administered at each span length. If the first two trials for a span were correct, the third trial was skipped, and if all trials for a span were incorrect, the task was discontinued. The dependent variable selected for analysis was a summary score calculated by first dividing the number of correct button presses by the total button presses for each span length and then summing these scores across all administered span lengths (including the one-item practice trials). Children had missing data for this task due to audiovisual malfunction ($n = 1$), examiner error ($n = 1$), and task noncompliance (e.g., disruptive temper tantrum or anxiety over touching computer screen that precluded completing the task, $n = 6$). The mean kappa value for interrater reliability was .99 (range = .76–1.00).

The Delayed Alternation task (adapted from Espy et al., 1999; Goldman, Rosvold, Vest, & Galkin, 1971) required children to retrieve a small reward from a well in a testing board. Each well was covered with an identical, neutrally colored cup. After a correct retrieval, the reward alternated to the opposite well. Thus, children needed to remember the previously rewarded location so as to maximize rewards. Rewards were hidden out of children's sight during the 10-s delay between trials, with the examiner also using a series of verbal distracters to break children's attentional focus during this delay. After an initial training phase where children needed to reach a criterion of three consecutive correct responses, up to 16 test trials were administered. To minimize battery length, if children made nine consecutive correct responses, the task was discontinued and children were given credit for any remaining trials because it was empirically determined from previous data (Espy et al., 1999) that children rarely made errors after eight consecutive correct responses. The dependent variable selected for analysis was the maximum number of consecutive correct responses minus the maximum number of consecutive incorrect responses; this measure conceptually represented children's ability to consistently maintain the task demands over the course of task administration. Although this variable was preferred conceptually, it was kurtotic (kurtosis = 4.83); thus, the outliers were trimmed to the value 3 standard deviations above the mean, which brought the distributional properties into the acceptable range (kurtosis < 3) prior to analysis. One child had missing data because of task noncompliance. Mean interrater reliability was 99.8% (range = 94–100).

Four tasks—Big–Little Stroop, Go/No-Go, Shape School (Inhibit condition), and Snack Delay—were postulated to place stronger demands on inhibitory control. In the Big–Little Stroop (adapted from Kochanska et al., 2000), administered via computer, children were shown line drawings of everyday objects containing smaller embedded pictures that either matched or conflicted in identity with these objects. To prime the salience of the large shape, each trial was preceded by a brief presentation (730 ms) of the large shape. Children were required to name the smaller embedded pictures. After a pretest to ensure that children could name each picture in the stimulus set, 24 trials were administered (50% conflict trials). Children who were unable to correctly name the majority of the stimuli in the pretest were assigned a score of 0 ($n = 3$). The proportion of correct responses on the conflict trials was selected as the dependent variable. Children had missing data on this task due to audiovisual malfunction ($n = 2$) and task noncompliance ($n = 4$). The mean interrater reliability kappa value was .91 (range = .77–.98).

In the Go/No-Go task (adapted from Simpson & Riggs, 2006), children were presented with pictures of colored fish and asked to "catch" these fish by responding on a button box. On less frequent "no-go" trials (25% of trials), a stimulus image of a shark appeared and children were instructed to "let it go" by withholding the button press response. Feedback was provided in the form of a fishing net, which broke when children made an error of commission by pressing the button in response to the shark stimulus. The "go" and "no-go" stimuli were presented for up to 1500 ms, with an interstimulus interval of 1000 ms. The dependent variable for this task was the d prime (d') measure (i.e., the standardized difference between the hit rate and the false alarm rate, calculated by subtracting the z-score value of the hit rate right-tail p value from the z-score value of the false alarm rate right-tail p value). The d' statistic is used routinely in the signal detection literature and reflects better discrimination of two classes of stimuli. Because this index integrates accuracy for both "go" and "no-go" trials, data for children with relatively poor attentional deployment (e.g., many misses on "go" trials) can be used, making this measure advantageous in young preschoolers. Children had missing data due to examiner error ($n = 1$) and task noncompliance ($n = 6$).

The Inhibit condition of the Shape School task (adapted from Espy, 1997) required children to name the color of a cartoon stimulus when the stimulus had a happy face and to suppress this naming response and remain silent when the stimulus had a sad face. For this study, stimuli were digitized and presented via computer instead of the previously used storybook format. First, a set of 12 control condition trials was administered, during which children named the colors of stimuli with neutral faces to prime the prepotent color-naming response. For the Inhibit condition, children were taught the "inhibit rule" through the story format and a set of six initial trials were administered. If children could not perform these initial trials, task administration was discontinued. Those children who were unable to complete the task to this point or who failed these initial trials were assigned a score of 0 ($n = 44$). Children who completed the initial trials were then administered the 18 Inhibit condition test trials (12 color naming and 6 naming suppression). The percentage of correct suppression responses was selected as the dependent variable. Children had missing data due to audiovisual malfunction ($n = 11$), examiner error ($n = 1$), and task noncompliance ($n = 21$). The mean interrater reliability kappa value was .92 (range = .80–.98).

In the Snack Delay task (adapted from Kochanska et al., 1996; Korkman, Kirk, & Kemp, 1998), a handful of M & M candies was placed under a transparent cup directly in front of participants. Children were instructed to stand still with their hands on a placemat decorated with two handprints, without moving or talking, until the examiner rang a bell signaling the end of the task when children could eat the candies. During the 240-s delay, the examiner initiated a series of distracters (e.g., coughing, dropping her pencil), and near the conclusion of the task, she left the room for a period of 90 s. The selected dependent measure was a summary score, where children were given up to 3 points for each 5-s epoch (1 point for standing still, 1 point for keeping their hands on the mat, and 1 point for remaining silent) summed across all epochs prior to children eating the snack or until the task ended at 240 s. Children had missing data due to audiovisual malfunction ($n = 1$) and task noncompliance ($n = 13$). Mean interrater agreement was 89% (range = 81–95).

Statistical methods

Descriptive statistics were calculated using SAS 9.2 (SAS Institute, Cary, NC, USA), with CFA analyses conducted in Mplus 6 (Muthén & Muthén, 2006). Across the executive tasks, 86% to 100% of data were available. In many cases, missing data seemed unlikely to be systematically related to EF (e.g., examiner error, technical difficulties), but it is not possible to rule out the possibility that a subset of the missing data might not be missing at random (MAR). However, in a simulation study, when the proportion of missing data was low and the cause of "missingness" was moderately related to the missing data, violation of the MAR assumption had negligible effects on outcomes (Collins, Schafer, & Kam, 2001). Therefore, although it was not possible to directly test whether the MAR assumption held for this dataset, the potential impact of a small proportion of data that was not missing at random

(NMAR) was deemed minimal. Missing data were estimated on the basis of all available data points using the EM algorithm in Mplus (Muthén & Muthén, 2006).

To evaluate latent EF structure, several models of task performance were considered based on the developmental and neuropsychological literature as well as analyses of nonexecutive task features and demands. First, a unitary model was tested, where all tasks loaded on a single common factor. Next, a two-factor model that consisted of both working memory and inhibition constructs was tested. Finally, additional alternative models were developed where tasks were grouped based on task characteristics and nonexecutive demands, including modality of administration (computerized or not), presence or absence of ongoing feedback about performance accuracy, whether spatial information was integral to the task, the degree of language demands, and whether children were performing the task under timed conditions. (Specific task assignment to the latent factors in each model are provided in the Results section.) Model fit to the data was assessed using several fit statistics. The chi-square (χ^2) statistic provided a global indication of model fit, where a nonsignificant χ^2 value suggests adequate model fit (Brown, 2006; Schumaker & Lomax, 2004). Because the χ^2 test is overly sensitive to deviations from perfect fit in large samples, additional indexes used for model evaluation and comparison included the root mean square error of approximation (*RMSEA*) (Browne & Cudeck, 1993), the comparative fit index (*CFI*), and the Bayesian information criterion (*BIC*). *RMSEA* values less than .06 and *CFI* indexes between .95 and 1.00 indicate good fit (Hu & Bentler, 1999; Yu, 2002). All models were nested, so the χ^2 difference test was used to compare model fit. Where models do not differ significantly, with a p value less than .05, the simpler model is preferred on the basis of parsimony (Bollen, 1989). The *BIC* can also be used to compare models; a difference in *BIC* values greater than 10 points indicates that the model with the smaller *BIC* value provides the better fit (Raftery, 1993).

After a best-fitting model was chosen, factorial invariance of the measurement model was evaluated by dividing the total sample into groups by sex and then by sociodemographic risk (low risk vs. at risk). In a series of increasingly restrictive nested models, parameters were constrained to equality across groups using model fit statistics and χ^2 difference tests to determine whether the same measurement model, with the same factors, loadings, intercepts, and/or residual variances, applies to both subgroups (Meredith, 1993; Meredith & Horn, 2001). When the measurement model is invariant, latent factor parameters can be compared between groups, including the latent mean that reflects the average score on the latent construct for each group.

Results

Descriptive statistics for all of the dependent variables considered for inclusion in the CFA are presented in Table 1. Dependent variables were selected as indicators for use in the CFA only if they displayed adequate distributional characteristics without substantial skewness or kurtosis. For the dependent variables selected as indicators for each executive task, descriptive statistics for boys and girls, and for at-risk and low-risk children, are presented in Table 2, with tests of the significance of sex and sociodemographic risk differences. Boys and girls did not differ in manifest task performance on any of the dependent variables selected for analysis. When examining relations between sociodemographic risk and manifest task performance, the low-risk children performed significantly better on the selected dependent variables for five tasks: Nebraska Barnyard, Big–Little Stroop, Go/No-Go, Shape School, and Snack Delay. Separate analyses testing the interactions between sex and sociodemographic risk were also conducted, and there were no statistically significant interactions. Task intercorrelations among the variables selected for analysis were low to moderate in magnitude (see Table 3).

For each model, fit statistics and χ^2 difference tests were used to determine which of the proposed models best fit 3-year-old executive task performance (see Table 4). The unitary model, where all tasks loaded on a single common factor (illustrated in Fig. 1, Model 1), fit the obtained data well, with a nonsignificant χ^2 value and adequate fit statistics. Next, the model that separated inhibition from working memory demands was estimated (Fig. 1, Model 2). This model also showed adequate fit to the data but did not result in a significant improvement in fit over the unitary model, so the unitary model was retained as more parsimonious. The correlation between the factors was high (r = .76), sug-

Table 2
Descriptive statistics for selected EF indicators by sex and sociodemographic risk group.

Task and dependent measure	Boys (n = 96–113)		Girls (n = 99–115)			At-risk (n = 75–96)		Low-risk (n = 120–132)		
	M	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
Nine Boxes: Longest correct run	4.50	1.52	4.65	1.65	0.75	4.50	1.71	4.63	1.49	0.59
Nebraska Barnyard: Summary score	1.12	0.57	1.21	0.64	1.05	0.96	0.50	1.30	0.64	4.41***
Delayed Alternation: Longest correct run–longest incorrect run	−0.68	3.40	−0.07	3.81	1.28	−0.59	3.63	−0.22	3.61	0.75
Big–Little Stroop: Proportion correct (conflict trials)	0.31	0.30	0.26	0.29	−1.25	0.21	0.24	0.35	0.31	3.75**
Go/No-Go: *d'*	0.55	0.47	0.60	0.56	0.73	0.47	0.35	0.66	0.60	2.86*
Shape School: Proportion correct (suppression stimuli)	0.33	0.40	0.39	0.43	1.11	0.28	0.38	0.40	0.43	1.98*
Snack Delay: Summary score	31.94	24.73	38.58	23.28	2.02*	31.34	21.40	38.05	25.63	2.07*

* $p < .05$.
** $p < .001$.
*** $p < .0001$.

Table 3
Correlations among selected EF indicators.

Task	1	2	3	4	5	6
1. Nine Boxes	–					
2. Nebraska Barnyard	.19*	–				
3. Delayed Alternation	.10	.17*	–			
4. Big–Little Stroop	.08	.28**	.23**	--		
5. Go/No-Go	.04	.18*	.17*	.24**	–	
6. Shape School	.03	.30**	.16*	.40**	.26**	–
7. Snack Delay	.12	.19*	-.03	.25**	.18*	.25**

* $p < .05$.
** $p < .001$.

Table 4
Goodness-of-fit indexes for alternative CFA models of preschool EF.

Model and number of factors[a]	χ^2	*df*	*p*	RMSEA	CFI	BIC	Model comparison	χ^2 difference	*df* difference	*p*
1. Unitary executive function (1)	14.84	14	.39	.02	.99	5074.01				
2. Working memory and inhibition (2)	12.40	13	.49	.00	1.00	5077.01	*Model 1* vs. Model 2	2.44	1	.12
3. Spatial and nonspatial tasks (2)	14.58	13	.33	.02	.99	5079.183	*Model 1* vs. Model 3	0.26	1	.61
4. Feedback and nonfeedback tasks (2)	14.23	13	.36	.02	.99	5078.83	*Model 1* vs. Model 4	0.61	1	.43

Note. For model comparisons, the favored model is italicized. When two models did not differ statistically, the more parsimonious model was chosen (Bollen, 1989).
[a] Numbers of factors are in parentheses.

gesting that the two factors had little unique explanatory power. Because the two-factor working memory and inhibition model was developed based on theory, we examined the variance accounted for by the latent factors in this model relative to the unitary EF model. In general, the differences in task variance accounted for across the two models were small (r^2 difference = 0–2.9%) with the exception of Noisy Book, where 13.5% more variance was explained in the two-factor model than in the

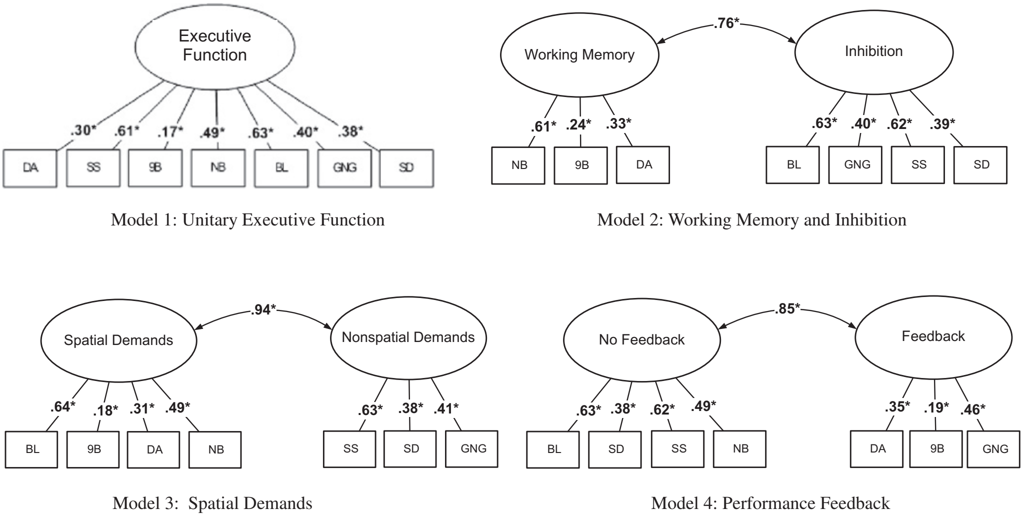

Fig. 1. Alternative CFA models of preschool EF. 9B, Nine Boxes task; BL, Big–Little Stroop; DA, Delayed Alternation task; GNG, Go/No-Go task; NB, Nebraska Barnyard task; SD, Snack Delay task; SS, Shape School task (Inhibit condition). Standardized factor loadings and coefficients are shown.

unitary model. Because the overall model fit did not differ between models, and only one task showed some evidence of specificity for measuring working memory, this evidence was not deemed compelling to favor choosing the two-factor model over the more parsimonious unitary model.

Next, a number of models were tested where EF tasks were grouped according to their nonexecutive features or demands. Of five models tested, only two could be fit successfully to the data: a model grouping tasks as to whether the tasks had higher or lower spatial demands and a model grouping tasks according to whether children were provided with ongoing feedback about their performance (Fig. 1, Models 3 and 4). Again, model fit was adequate, but the correlation between the two latent factors was very high (rs = .94 and .85) and fit was not a significant improvement over the unitary factor. Thus, even in young preschoolers, the other nonexecutive task demands measured by these latent factors were not the primary construct underlying variability in task performance. Three additional models were constructed but could not be estimated: a model grouping timed and untimed tasks, a model grouping tasks administered via computer and noncomputerized tasks, and a model grouping tasks with higher and lower language demands. In each case, an estimated interfactor correlation greater than 1.0 resulted in a nonpositive definite matrix. Thus, consistent with the findings of Wiebe and colleagues (2008) in an independent sample and with a different battery of tasks, the simplest single-factor model of EF was supported as sufficient to most parsimoniously account for the data.

The next step was to examine whether the same EF measurement model was appropriate for subsamples of preschoolers grouped by sex and sociodemographic risk group status. Invariance testing was conducted for the best-fitting unitary EF model only, and analyses are summarized in Table 5. For sex, there was support for metric invariance of the EF model. That is, factor loadings of each task on the unitary EF factor could be constrained to be equal in boys and girls. There was also support for scalar invariance, meaning that all of the intercepts for all of the indicators (in this case equivalent to the estimated indicator means) could be constrained to equality between boys and girls. Finally, models constraining error variances for the indicators and latent factor variances and means to be the same for boys and girls also were supported. Overall, the unitary EF model was determined to be equally well fitting for both boys and girls, but in contrast to the findings of Wiebe and colleagues (2008), there were no significant differences by sex in the latent variance or mean of EF or in the estimated means of the individual tasks.

Measurement invariance testing for sociodemographic risk grouping revealed a different pattern. Both metric and scalar invariance models were retained, meaning that the same task-to-factor

Table 5
Tests of measurement invariance of the unitary EF model for groups of children defined by sex (top panel) and sociodemographic risk status (bottom panel).

Model	χ^2	*df*	*p*	*RMSEA*	*CFI*	*BIC*	χ^2 difference	*df*	*p*
Sex invariance									
Baseline	29.95	28	.37	.03	.98	5165.15	–	–	–
Metric invariance	35.30	34	.41	.02	.99	5137.93	5.35	6	.50
Scalar invariance	46.68	40	.22	.04	.94	5116.73	11.38	6	.08
Equal error variances	51.43	47	.30	.03	.96	5083.47	4.75	7	.69
Equal factor variances	52.00	48	.32	.03	.97	5078.62	0.57	1	.45
Equal factor means	52.83	49	.33	.03	.97	5074.01	0.83	1	.36
Sociodemographic risk status invariance									
Baseline	26.05	28	.57	.00	1.00	5107.31	–	–	–
Metric invariance	34.88	34	.43	.02	.99	5083.56	8.83	6	.18
Scalar invariance	41.25	40	.42	.02	.99	5057.37	6.37	6	.38
Equal error variances	73.80	47	.01	.07	.71	5051.90	32.55	7	<.0001
Partial error variances[a]	47.48	45	.37	.02	.97	5036.44	6.23	5	.28
Equal factor variances[a]	57.36	46	.12	.05	.88	5040.89	9.88	1	.002
Equal factor means[a]	79.55	47	.002	.08	.65	5057.65	22.19	1	<.0001

[a] Nebraska Barnyard and Go/No-Go error variances permitted to differ between low-risk and at-risk groups.

loadings and intercepts (or estimated indicator means) for all tasks could be retained for low-risk and at-risk groups. However, error variances of the indicators could not be constrained to equality across sociodemographic risk groups without a significant worsening of model fit to the data (see Table 5). To explore the sources of variance in measurement, a partial invariance model was tested by constraining each indicator error variance to equality individually for each EF task indicator in turn (Byrne, Shavelson, & Muthen, 1989; Meade & Bauer, 2007). A partial invariance model for indicator error variances was fit to the data, where the intercepts for the Go/No-Go and Nebraska Barnyard tasks were allowed to differ in low-risk and at-risk groups and all other task error variances were set to equality. For both tasks, error variance estimates were higher for the low-risk group than for the at-risk group (Go/No-Go εs = .300 and .109, and Nebraska Barnyard εs = .320 and .202, respectively). Successive invariance constraints could then be added to this model. Follow-up models constraining factor variances and factor means to be equal for low-risk and at-risk groups also were not supported. The latent EF variance was higher for the low-risk group (1.42) than for the at-risk group (0.48), and mean latent EF was 1.19 standard deviations higher for low-risk children relative to their at-risk peers. The final best-fitting model for sociodemographic risk is illustrated in Fig. 2.

Discussion

The goal of this study was to examine the structure of EF in preschoolers at 3 years of age. A large sample of children from diverse sociodemographic backgrounds was assessed within weeks of their

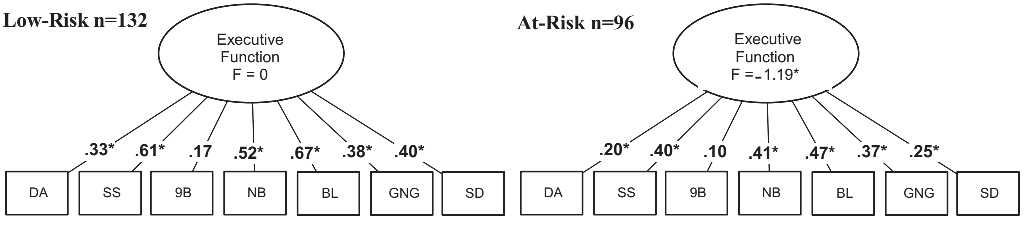

Fig. 2. Preferred model of unitary preschool EF in subgroups of children defined by sociodemographic risk group status. Go/No-Go and Nebraska Barnyard error variances, factor variances, and factor means are allowed to differ between low-risk and at-risk groups. 9B, Nine Boxes task; BL, Big–Little Stroop; DA, Delayed Alternation task; GNG, Go/No-Go task; NB, Nebraska Barnyard task; SD, Snack Delay task; SS, Shape School task (Inhibit condition). Standardized factor loadings and coefficients are shown.

third birthday using age-appropriate tasks selected to vary in both executive and nonexecutive demands. This study was designed to empirically evaluate the fit of alternative models of EF structure at an important transitional age that marks the conclusion of the toddler years and entry into preschool and immediately precedes substantial gains in executive task performance. Although tasks were chosen as putative measures of inhibitory control and working memory processes, the two-factor model of EF did not have significantly more explanatory power than the simplest unitary EF model. For all dual-factor models that could be estimated, the correlation between factors was high ($r \geqslant .75$), suggesting substantial overlap between the factors. This finding replicates that of Wiebe and colleagues (2008), but in a better ascertained sample of children assessed within a narrow age window at the beginning of the preschool period and with more complete data on a new battery of carefully selected, age-appropriate tasks. The unitary EF factor structure was chosen based on relative and absolute model fit. However, this conclusion by no means indicates that this issue is settled. Rather, it is possible that the two-factor structure might better predict variation in important outcomes such as externalizing symptomatology or academic skills, or there might be latent clusters of preschoolers who show systematically differing variations in working memory and inhibition skills that, when combined, result in a single factor as the best fitting. In the absence of such evidence, the current findings best support a unitary conceptualization of EF early during the preschool period.

The unitary EF model was tested to determine whether the same model held equally for boys and girls. Consistent with the findings of Hughes and colleagues (2010), and in contrast to the findings of Wiebe and colleagues (2008), boys and girls did not differ in latent EF. This inconsistency could be attributable in part to other domains; for example, there are pronounced differences in language development rates during early childhood, and a number of studies have linked verbal proficiency to executive task performance (Hughes, 1998; Wolfe & Bell, 2003). Because the executive battery included in the current study included tasks varying in verbal demands, it may have been better suited to separating true EF variance from nonexecutive variance attributable to linguistic requirements of some tasks.

Parallel analyses were conducted for groups of children differing in sociodemographic risk. The finding that at-risk children had lower levels of EF than low-risk children is consistent with other research. In their longitudinal study, Hughes and colleagues (2010) found that family income was related to children's initial level of latent EF at 4 years of age but was not associated with the rate of change across the 2-year period. In contrast, Wiebe and colleagues (2008) found no relation between maternal education and latent EF. In the current study, stratified sampling methods were used to increase power to test relations between EF structure and sociodemographic risk; furthermore, eligibility for public health insurance is likely a better proxy of a family's risk status than maternal education. Other studies have shown relations between socioeconomic status and performance on individual EF tasks (Noble et al., 2005; Noble et al., 2007) in school-age children. Because of the young age (3 years) of children in the current study, more children of this age spend a greater proportion of their day in the home environment and are less likely to participate in, and spend less time in, organized preschool programs. As a result, one might expect that effects of differing sociodemographic characteristics of the family on EF would be greater at this young age. In addition to differences in latent EF means and variances, there were differences in error variances for two of the indicators, the Go/No-Go and Nebraska Barnyard tasks. Essentially, unequal error variance indicates that these particular tasks are not equally precise measures of EF for at-risk and low-risk groups of children. However, this step of invariance testing is generally considered to be less important than metric or scalar invariance; indeed, some dispense with this step as being too restrictive (Brown, 2006; Little, 1997). Thus, these differences between risk groups do not preclude interpreting differences in latent EF.

Studying the emergence of EF during the early preschool years is challenging. However, even at 36 months of age, it was possible to select a dependent measure for each task that displayed appropriate distributional properties. In our approach, we aimed to balance conceptual issues (e.g., face validity of chosen measure), sensitivity across the full range of performance, and psychometric issues (e.g., avoiding pass/fail measures of performance). Psychometric issues are particularly salient during early childhood because reliability of any test is typically lower than would be expected in older children or adults. Reliability reflects meaningful variability in performance and, consequently, places an upper limit on the relation between the manifest variable and latent EF. Other influences that reduce task reliability might be systematic and meaningful individual differences between children (e.g., language ability, activity level) as well as variability of less theoretical interest (e.g., day-to-day variation in children's alertness, familiarity with computers). This issue is precisely what CFA is designed to address and highlights the benefit of interpreting latent factor scores rather than individual task performance, particularly in the preschool age range. The finding that the proportion of error variance (and hence the "true" EF variability) reflected in Go/No-Go and Nebraska Barnyard scores differed by sociodemographic risk speaks to this point. Both of these tasks required children to interact directly with the computer (by using a button box or touch screen). In 3-year-olds, this requirement may have introduced additional nonexecutive sources of variation. The remaining computer-based tasks in the battery required children to produce a verbal response (naming pictures in Big–Little Stroop and naming colors in Shape School task) for which all children likely had sufficient experience. Notably, in studies where children's manifest performance on executive tasks is compared directly, there is an implicit assumption of measurement invariance, that is, that tasks are measuring the same EF construct in the same way across groups that are being compared. If this assumption is not valid, the findings are not necessarily robust.

In comparing the findings of this and other recent preschool studies with the adult literature, it is important to note that CFA studies of adult EF structure have shown that the three main components of EF share a great deal of variance, indicating some degree of unity even during adulthood. The precise nature of this common "core" largely remains to be determined. Miyake and colleagues (2000) suggested that it might be related to representational inhibition and/or working memory maintenance. Friedman and colleagues (2008) found that the heritable component of variation in EF almost completely overlapped with the inhibition factor but that working memory updating and set shifting each had a unique additional component. Friedman and colleagues speculated that the inhibition factor might in fact reflect goal activation, which is required across all EF tasks. Friedman, Miyake, and their colleagues did not discuss development explicitly, but it is possible that preschool unitary EF is largely similar to the core EF component during adulthood, and development entails the emergence of additional cognitive processes that contribute to working memory and shifting ability. This account is similar to the model proposed by Garon and colleagues (2008), although in this model working memory emerges first, followed by inhibition and then shifting, driven by growth in the ability to deploy and coordinate attention. To address this question, further work is necessary to trace the development of EF structure beyond the preschool years.

Although statistical modeling using CFA has proven to be an effective tool to reveal cognitive organization, the obtained results complement others using traditional within-task analyses. To conduct a CFA, a single outcome variable should be selected from each task because multiple variables from a single task are typically more highly correlated with each other than are indicators from different tasks due to shared method variance (Gorsuch, 1983). This strategy, however, limits the ability to examine the impact of within-task manipulations on performance. For example, the Shape School task includes several control conditions and also blocks adding set-shifting requirements (Espy, 1997; Espy, Bull, Martin, & Stroup, 2006). Further comparison of contrasting performance within graduated measures such as this will help to determine the extent to which differences in performance on these putative EF measures are better accounted for by differences in underlying cognitive abilities. Furthermore, in the current study, sociodemographic risk is only a proxy for a variety of environmental factors. Proximal factors that directly affect children's development include parenting style, environmental stimulation within the home and neighborhood, and community resources, among many other factors. Further work with this sample analyzing the relations between specific contextual factors and child outcomes is planned and will provide a more nuanced understanding of differences in EF observed in the current study between low-risk and at-risk children.

Despite these limitations, this study adds to a growing body of evidence that EF measured during the preschool years is simpler in form than EF measured later in development (Hughes et al., 2010; Wiebe et al., 2008). Furthermore, given that EF has been found to be separable into three component processes in samples 7 years of age or older (Huizinga et al., 2006; Lehto et al., 2003), the current findings add to the claim that the preschool period is marked by tremendous change in EF. Not only does EF gain in efficiency over this period, as has been demonstrated previously (e.g., Carlson, 2005), but also fundamental structural changes must occur. A similar approach to assessment at later time points is now required to understand how EFs might progressively differentiate during the preschool years. The children in the current study have been followed longitudinally, completing 3 additional follow-up assessments over the following 2.25 years of development. Children complete the same task battery at all ages, so that it will be possible to directly test whether the same model continues to adequately account for data or whether a more complex differentiated model is necessary to describe EF as children mature. Delineating the developmental trajectories of latent EF likely will reveal fundamental insights into how variables that here were shown to be associated with differences in initial EF status (i.e., sex and sociodemographic risk) affect the dynamic unfolding of self-regulatory capacity over the preschool years.

Acknowledgments

This work was supported by National Institutes of Health (NIH) Grants MH065668 and DA014661 to Kimberly Andrews Espy, Grant DA024769 to Sandra Wiebe, and Grant DA023653 to Kimberly Andrews Espy and Lauren Wakschlag. We thank Alison Brindle, Colleen McLaughlin, Doran Hadan, Rachel Hoien, Rachel Huggins, Matthew Moehr, and other members of the Developmental Cognitive Neuroscience Laboratory for assistance with data collection and coding.

References

Allhusen, V., Belsky, J., Booth-LaForce, C., Bradley, R., Brownell, C., Burchinal, M., et al (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development, 76*, 795–810.

Assel, M. A., Landry, S. H., Swank, P., Smith, K. E., & Steelman, L. M. (2003). Precursors to mathematical skills: Examining the roles of visual–spatial skills, executive processes, and parenting factors. *Applied Developmental Science, 7*, 27–38.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Brocki, K. C., Nyberg, L., Thorell, L. B., & Bohlin, G. (2007). Early concurrent and longitudinal symptoms of ADHD and ODD: Relations to different types of inhibitory control and working memory. *Journal of Child Psychology and Psychiatry, 48*, 1033–1041.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement. *Developmental Neuropsychology, 33*, 205–228.

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T., & Nelson, J. M. (in press). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent mathematic skills. *Developmental Science.*

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–616.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology, 8*, 219–228.

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science, 11*, 339–353.

Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351.

Deák, G. O., Rey, S. D., & Pick, A. D. (2004). Effects of age, reminders, and task difficulty on young children's rule-switching flexibility. *Cognitive Development, 19*, 385–400.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control of reaching. *Annals of the New York Academy of Sciences, 608*, 637–676.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

Diamond, A., Briand, L., Fossella, J., & Gehlbach, L. (2004). Genetic and neurochemical modulation of prefrontal cognitive functions in children. *American Journal of Psychiatry, 161*, 125–132.

Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development, 62*(4), 1–206.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "do as I say, not as I do". *Developmental Psychobiology, 29*, 315–334.

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology, 26*, 379–384.

Espy, K. A., Bull, R., Martin, J., & Stroup, W. (2006). Measuring the development of executive control with the Shape School. *Psychological Assessment, 18*, 373–381.

Espy, K.A., Sheffield, T.S., Wiebe, S.A., Clark, C.A., & Moehr, M. (in press). Executive control and dimensions of problem behaviors in preschool children. *Journal of Consulting and Clinical Psychology*.
Espy, K. A., Kaufmann, P. M., McDiarmid, M. D., & Glisky, M. L. (1999). Executive functioning in preschool children: Performance on A-not-B and other delayed response formats. *Brain and Cognition, 41*, 178–199.
Farah, M. J., Betancourt, L., Shera, D. M., Savage, J. H., Giannetta, J. M., Brodsky, N. L., et al (2008). Environmental stimulation, parental nurturance, and cognitive development in humans. *Developmental Science, 11*, 793–801.
Fassbender, C., Murphy, K., Foxe, J. J., Wylie, G. R., Javitt, D. C., Robertson, I. H., et al (2004). A topography of executive functions and their interactions revealed by functional magnetic resonance imaging. *Cognitive Brain Research, 20*, 132–143.
Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., et al (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18*, 893–900.
Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*, 101–135.
Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201–225.
Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. *NeuroImage, 17*, 1820–1829.
Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Review, 134*, 31–60.
Goldman, P. S., Rosvold, H. E., Vest, B., & Galkin, T. W. (1971). Analysis of the delayed alternation deficit produced by dorsolateral prefrontal lesions in the rhesus monkey. *Journal of Comparative and Physiological Psychology, 77*, 212–220.
Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Hanania, R. (2010). Two types of perseveration in the Dimensional Change Card Sort. *Journal of Experimental Child Psychology, 107*, 325–336.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253.
Hughes, C. (2002). Executive functions and development: Why the interest? *Infant and Child Development, 11*, 69–71.
Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *Journal of Child Psychology and Psychiatry, 39*, 981–994.
Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology, 28*, 645–668.
Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20–36.
Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*, 2017–2036.
Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28*, 517–527.
Johnson, M. H. (1995). The inhibition of automatic saccades in early infancy. *Developmental Psychobiology, 28*, 281–291.
Kochanska, G., Murray, K., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity, change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220–232.
Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507.
Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: Psychological Corporation.
Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59–80.
Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews, 30*, 718–729.
Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.
Magnuson, K. A., Sexton, H. R., Davis-Kean, P. E., & Huston, A. C. (2009). Increases in maternal education and young children's language skills. *Merrill–Palmer Quarterly, 55*, 319–350.
Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*, 689–704.
McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*, 222–243.
McNab, F., Leroux, G., Strand, F., Thorell, L., Bergman, S., & Klingberg, T. (2008). Common and unique components of inhibition and working memory: An fMRI, within-subjects investigation. *Neuropsychologia, 46*, 2668–2682.
Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling, 14*, 611–635.
Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 525–543.
Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development, 75*, 1373–1386.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide* (4th ed.). Los Angeles: Authors.

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science, 10*, 464–480.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*, 74–87.

Olson, S. L., Sameroff, A. J., Kerr, D. C. R., Lopez, N. L., & Wellman, H. M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Development and Psychopathology, 17*, 25–45.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., et al (1999). Structural maturation of neural pathways in children and adolescents: In vivo study. *Science, 283*, 1908–1911.

Prevor, M. B., & Diamond, A. (2005). Color–object interference in young children: A Stroop effect in children 3½–6½ years old. *Cognitive Development, 20*, 256–278.

Raaijmakers, M. A. J., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., et al (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology, 36*, 1097–1107.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 163–180). Newbury Park, CA: Sage.

Schumaker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a "button-press" Go/No-Go task. *Journal of Experimental Child Psychology, 94*, 18–26.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience, 2*, 859–862.

Taylor, H. G., Minich, N. M., Klein, N., & Hack, M. (2004). Longitudinal outcomes of very low birth weight: Neuropsychological findings. *Journal of the International Neuropsychological Society, 10*, 149–163.

Thatcher, R. W. (1992). Cyclical cortical reorganization during early childhood. *Brain and Cognition, 20*, 24–50.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*, 1097–1118.

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language, 108*, 175–183.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure.. *Developmental Psychology, 44*, 575–587.

Wiebe, S. A., Espy, K. A., Stopp, C., Stewart, P., Respass, J., Jameson, T. R., et al (2009). Gene–environment interactions across development: Exploring DRD2 genotype and prenatal smoking effects on self-regulation. *Developmental Psychology, 45*, 31–44.

Wolfe, C. D., & Bell, M. A. (2003). Working memory and inhibitory control in early childhood: Contributions from physiology, temperament, and language. *Developmental Psychobiology, 44*, 68–83.

Yu, C-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Doctoral dissertation, University of California, Los Angeles.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.