Journal of Cognition and Development

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hjcd20

A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children

Michael R. Miller [a] , Gerald F. Giesbrecht [a] , Ulrich Müller [a] , Robert J. McInerney [a] & Kimberly A. Kerns [a]

[a] University of Victoria , Canada

To cite this article: Michael R. Miller , Gerald F. Giesbrecht , Ulrich Müller , Robert J. McInerney & Kimberly A. Kerns (2012) A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children, Journal of Cognition and Development, 13:3, 395-423, DOI: 10.1080/15248372.2011.585478

To link to this article: http://dx.doi.org/10.1080/15248372.2011.585478

JOURNAL OF COGNITION AND DEVELOPMENT, 13(3):395–423
Copyright © 2012 Taylor & Francis Group, LLC
ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2011.585478

Psychology Press
Taylor & Francis Group

A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children

Michael R. Miller, Gerald F. Giesbrecht, Ulrich Müller,
Robert J. McInerney, and Kimberly A. Kerns
University of Victoria, Canada

The composition of executive function (EF) in preschool children was examined using confirmatory factor analysis (CFA). A sample of 129 children between 3 and 5 years of age completed a battery of EF tasks. Using performance indicators of working memory and inhibition similar to previous CFA studies with preschoolers, we replicated a unitary EF factor structure. Next, additional performance indicators were included to distinctly measure working memory, set shifting, and inhibition factors. A two-factor model consisting of working memory and inhibition fit the data better than both a single-factor model and a three-factor model. Findings suggest that the structure of EF in preschoolers that emerges from CFA is influenced by task and performance indicator selection.

Executive function (EF) refers to higher mental processes that are involved in the conscious control of actions and thoughts (Zelazo & Müller, 2010). Historically, the construct of EF has been used to refer to psychological processes the impairment of which are presumed to underlie manifest behavioral deficits in patients with lesions to the prefrontal cortex (Fuster, 1989). In the last two decades, EF has received considerable attention in

---

Michael R. Miller is now at Vanderbilt University, Nashville, TN, USA. Gerald F. Giesbrecht is now at the Behavioral Research Unit, Alberta Children's Hospital, Calgary, AB, Canada. Robert J. McInerney is now at IWK Health Centre, Halifax, NS, Canada.

Correspondence should be sent to Michael R. Miller, Department of Psychology and Human Development, Vanderbilt University, Peabody College #552, 230 Appleton Place, Nashville, TN 37203-5721, USA. E-mail: m.miller@vanderbilt.edu

developmental psychology because it has been shown that: a) EF undergoes dramatic developmental changes that are particularly pronounced during the preschool period (Carlson, 2005; Diamond, 2006; Garon, Bryson, & Smith, 2008); b) EF is related to children's social understanding (Carlson & Moses, 2001; Flynn, 2007; Sabbagh, Xu, Carlson, Moses, & Lee, 2006) as well as their school readiness and achievement (Müller, Liebermann, Frye, & Zelazo, 2008); and c) impairments in EF are involved in different developmental disorders (Jurado & Rosselli, 2007; O'Hearn, Asato, Ordaz, & Luna, 2008; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005).

Even though there is general agreement about the importance of EF for adaptive functioning, the construct of EF has remained elusive (Jurado & Rosselli, 2007; Zelazo & Müller, 2010). There are two major issues in the conceptualization of EF that have prevented clear and definitive conclusions: the structure of EF and the composition of the tasks used to measure EF. The first issue concerns the question of whether EF is a unitary construct or a heterogeneous set of dissociable processes (Garon et al., 2008; Jurado & Rosselli, 2007). One common approach to address this issue has been to devise comprehensive neuropsychological test batteries and to use principal components analysis (PCA) or exploratory factor analysis (EFA) to determine whether the manifest variables can be reduced to a smaller number of underlying factors. Developmental studies using PCA and EFA have generally revealed between one and four factors of EF in preschool children (Carlson, Mandell, & Williams, 2004; Carlson & Moses, 2001; Espy, Kaufmann, McDiarmid, & Glisky, 1999; Hughes, 1998; Hughes & Ensor, 2007; Welsh, Nix, Blair, Bierman, & Nelson, 2010) and three factors in school-aged children (Brocki & Bohlin, 2004; Levin et al., 1991; Pennington, 1997; Welsh, Pennington, & Groisser, 1991; for a detailed summary, Zelazo & Müller, 2010). The factorial solutions derived in these studies differ in terms of the number, composition, and interpretation of the extracted factors, limiting the conclusions about the nature of EF that can be drawn from these studies. These inconsistencies may be due to the use of different test batteries and to the age ranges of the participants (van der Sluis, de Jong, & van der Leij, 2007). Furthermore, many PCA and EFA studies have used Varimax rotation, which restricts the factor solution to uncorrelated or orthogonal factors and tends to produce factors that are sample specific and difficult to replicate (Gorsuch, 1997; Wiebe, Espy, & Charak, 2008).

The second major issue that has created difficulties in clearly delineating EF is referred to as the *task impurity problem*. Task impurity is a problem for understanding the structure of EF because tasks designed to measure EF typically make more than one kind of executive processing demand (Hughes & Graham, 2002), and they also may involve a variety of nonexecutive processes (e.g., perceptual processing) that can influence task performance (Miyake et al., 2000).

An advanced methodological approach that can be used to address the issues associated with the conceptualization of EF is confirmatory factor analysis (CFA). In CFA, correlations are examined among unmeasured, latent variables that are composed of two or more observed, manifest variables. In contrast to PCA and EFA, in which the data determine the underlying factor model, CFA evaluates how well a theory-driven factor model fits the data and allows one to compare competing theory-driven models in terms of their fit to the data. Whereas the new component variables are functions of the original manifest variables in PCA, the manifest variables are functions of the underlying latent factors in EFA and CFA. As a result, both EFA and CFA increase the potential for reliable associations between a latent factor and its particular set of manifest variables by estimating and isolating unique sources of variance in the manifest variables (Bryant & Yarnold, 1994). However, unlike PCA and EFA, in which each manifest variable loads on every factor in the analysis, CFA allows one to specify the loadings for each latent factor to better satisfy a-priori hypotheses. Therefore, CFA is a powerful tool for evaluating different hypotheses about the structure of EF. Furthermore, CFA addresses the task impurity problem by extracting only the common variance shared by different EF tasks that are stipulated to measure the same latent factor; this results in a purer measure of the EF construct (Miyake et al., 2000).

Miyake and colleagues (2000) pioneered the use of CFA in the study of EF by identifying set shifting, updating or working memory, and response inhibition as three core components of EF in an adult sample. Each of these components was assessed by three simple EF tasks. CFA supported the three-factor model, with significant interrelations among all three latent variables. Miyake et al. (2000) interpreted these findings as suggesting that set shifting, working memory, and response inhibition are "separable but moderately correlated constructs, thus indicating both unity and diversity of executive functions" (p. 87). Recent studies also have applied CFA to school-aged children. For instance, Lehto, Juujärvi, Kooistra, and Pulkkinen (2003) found that a three-factor model with set shifting, working memory, and inhibition best fit the data in a sample of 8- to 13-year-olds. Two further studies (Huizinga, Dolan, & van der Molen, 2006; van der Sluis et al., 2007) identified only set shifting and working memory as unique factors in school-aged children; the measures of inhibition did not significantly load on a common latent factor after controlling for processing speed.

To date, there have been three applications of CFA to EF involving typically developing preschool children (but see Schoemaker et al., 2012). Wiebe and colleagues (2008) administered a battery of working memory and inhibition measures that differed in task format and response requirements to a sample of children aged between 2 and 6 years old. The results of the CFA indicated that a single-factor model provided the best fit to the data. A further CFA with data from a sample of 3-year-olds who completed several measures of working memory and inhibition also supported the validity of a unitary EF factor (Wiebe et al., 2011). Finally, using a longitudinal design, Hughes, Ensor, Wilson, and Graham (2010) administered three EF tasks (working memory, inhibitory control, and planning) and found that a single latent EF factor provided a good fit to the data for children at both 4 and 6 years of age. Taking into account the diverse nature of EF that has been found in school-aged children (Huizinga et al., 2006; Lehto et al., 2003; van der Sluis et al., 2007), the findings from these three CFA studies with preschoolers suggest that the structure of EF undergoes a dramatic change between early and middle childhood.

For theoretical and methodological reasons, however, a more complex EF structure in preschool children should not be ruled out. From a theoretical perspective, it remains unclear how the single latent EF factor should be interpreted. Wiebe et al. (2008) suggested that the single factor reflected the "enhancement of relevant stimulus–response relations to achieve goal-oriented executive control of thought and action" (p. 584). Furthermore, they speculated that this enhancement might be particularly salient in preschoolers because the correct stimulus–response connections are potentially weaker in preschoolers than in older children. By contrast, Garon et al. (2008) argued that EF is based on attentional processes, which in turn, are not well integrated in preschool children. As a consequence, they suggested that "EF tasks that vary in their dependence on different attentional processes would not be strongly correlated during the early preschool period" (Garon et al., 2008, p. 51). Similarly, from a Vygotskian perspective (Vygotsky & Luria, 1994; see Fernyhough, 2010), it might be expected that EF becomes increasingly integrated with the internalization of speech and semiotic mediation of elementary EF processes. By comparison, the core knowledge perspective (Spelke, 2003) implies a similar integration process in that language is stipulated to combine representations that are delivered from the modular information processing systems.

From a methodological perspective, task selection and the choice of performance indicators may have influenced the findings of the three CFA studies of EF in preschoolers. By including only one task of three different EF constructs (working memory, inhibitory control, and planning), Hughes and colleagues (2010) were unable to compare competing EF models and acknowledged that "our results do not challenge the fractionated models of EF" (p. 31). Wiebe and colleagues (2008, 2011), in turn, may not have found a distinct inhibition factor for preschoolers because the performance indicators that were selected to measure inhibition did not involve a strong prepotent response. For example, Wiebe et al. (2008) selected a delayed response task (Goldman, Rosvold, & Mishkin, 1970) as an indicator of inhibition even though it only involved a 10-second delay. More importantly, the delayed response task is conceptually almost indistinguishable from another task also administered by Wiebe et al. (2008), the delayed alternation task (Espy et al., 1999). However, Wiebe et al. (2008) considered the delayed alternation task a measure of working memory. Similarly, instead of selecting errors of commission (i.e., responding to a stimulus when a response is supposed to be withheld) in the Go/No-Go task as a performance indicator for inhibition, Wiebe et al. (2011) selected *d* prime, which indexes the accuracy both for "go" and for "no-go" trials and likely reflects working memory and inhibitory control demands. Furthermore, it should be noted that other more complex models of EF (e.g., a working memory and inhibition two-factor model) that were tested and rejected by Wiebe and colleagues (2008, 2011) fit the observed data just as well as a single-factor model. In other words, the single-factor models did not outperform all other tested models but were favored for reasons of parsimony.

Based on these theoretical and methodological concerns, the present study had three goals. Building on the seminal work by Wiebe and colleagues (2008), the first goal of the present study was to replicate the unitary factor structure that Wiebe et al. (2008) found in their preschool sample by using similar performance indicators of working memory and inhibition. The second goal of the present study was to examine whether the selection of performance indicators affects the latent EF factor structure in preschoolers. Specifically, we examined whether selecting performance indicators that more clearly separate working memory and inhibition demands would result in a more diverse EF factor structure than has been previously found in CFA studies with typically developing preschoolers (Hughes et al., 2010; Wiebe et al., 2008, 2011). In this manner, our study looked to provide a strong test of the unitary EF factor model. If EF in preschool children is best represented as a unitary construct, than the unitary factor structure should hold regardless of the selected performance indicators. By contrast, if different components of EF are separable in preschoolers, then this distinction should be representative of the selected performance indicators. Finally, previous applications of CFA to EF in preschool children have not included indicators of set shifting. For this reason, it is unknown whether set shifting constitutes a latent EF factor in preschool children. Thus, the third goal of the present study was to assess this issue and, more broadly, to explore whether the EF data produced by preschoolers would fit the three-factor structure that has been found both in school-aged children (Lehto et al., 2003) and in adults (Miyake et al., 2000).

METHOD

Participants

One hundred thirty-one children between the ages of 3 and 5 years were recruited from day cares, from preschools, and through community advertisements in a metropolitan area of Southwestern Canada. Two children were dropped from the sample because of concerns about major developmental delays. The final sample consisted of fifty-five 3-year-olds, sixty-four 4-year-olds, and ten 5-year-olds, for a total of 129 children (51 girls; $M_{age} = 4;2$; $SD_{age} = 7$ months; age range: 3;0–5;8). The majority of the sample was Caucasian (about 80%) and came from two-parent families ($n = 113$) with 2 or more children ($n = 97$). Using a 5-point Likert scale ranging from elementary school to holding a graduate degree, the median maternal and paternal education of the sample was some college or university education.

Procedure

The present study used a within-person, cross-sectional design. All children were tested individually on a battery of tasks designed to measure aspects of EF, self-regulation, social understanding, and school readiness; only the data on EF are reported in this article. The task battery was divided into two 45-minute test sessions. The order in which children completed the two test sessions was randomly assigned, and the interval between sessions was never longer than 2 weeks. In addition, a fixed task order was chosen for both test sessions to separate tasks of similar cognitive demand and facilitate comparisons between EF tasks (for further justification, see Carlson & Moses, 2001). Two trained experimenters administered the tasks in a day care, a preschool, or a university laboratory setting. All test sessions were videotaped, and children received two small gifts ($1 in value) for their participation.

Measures

Backward digit and backward word spans. The Backward Span tasks were classified as measures of working memory (Carlson, Moses, & Breton, 2002). The procedure for administering both tasks followed that of Davis and Pratt (1995). With the aid of a hand puppet, the experimenter stated strings of single-digit, nonsequential numbers for the Backward Digit Span and single-syllable, nonsemantically related words for the Backward Word Span. Children were asked to verbally repeat the sequences in reverse order. The tasks began with a two-digit or two-word practice trial in which corrective feedback was provided as needed. If children were unsuccessful after two repetitions of the practice trial, the task ended, and they were given a score of 0 ($n = 79$ for the Backward Digit Span, and $n = 60$ for the Backward Word Span). The majority of children who received a 0 score were 3-year-olds ($n = 47$ for the Backward Digit Span, and $n = 41$ for the Backward Word Span). Children who passed the practice trial received two trials each of two-, three-, and four-digit or word lengths. The tasks were discontinued when children made errors on both trials of a given length. Performance was measured with the highest digit- or word-length completed.[1]

Boxes task. The Boxes task (Kerns & McInerney, 2007) was a computerized, self-ordered search task designed to measure working memory. Children were instructed to find a jack-in-the-box while continuously keeping in mind boxes they had already searched. The task began with a practice trial in which children were presented with two boxes and then told that Jack was hiding in one of the boxes. Through the use of a touchscreen monitor, children were instructed to select the box in which they thought Jack was hiding. The first box that children selected was always empty, and the second box always contained Jack. After finding Jack in the second box, a positive-feedback screen appeared with Jack smiling. Children were informed that Jack was going to hide again but that he never hides in the same box; he always hides in a different box. Then the boxes reappeared, and children were asked to find Jack again. Corrective feedback was provided as needed.

Following practice, children received two trials each of two, three, four, and five boxes. The rules were repeated before the start of every trial, but corrective feedback was not provided once a trial began. During each trial, children could make within- and between-search errors. A within-search error was recorded when children selected the same empty box two or more times before finding Jack. For example, children committed a within-search error when, in a three-box trial, they selected Box A, Box B, Box A, before finding Jack in Box C. A between-search error was recorded when children selected a box in which they had previously found Jack earlier in the trial. For example, children committed a between-search error when, in a three-box trial, they found Jack in Box A and then, after the positive-feedback screen had been presented, selected Box A again. The task was discontinued when children made between-search errors on both trials of a given length. Performance was measured with a ratio of the number of correct searches (i.e., hits) to the number of total searches (i.e., hits plus errors)

---

[1] Similar results emerged when performance was measured in terms of number of correct trials.

for completed three- and four-box trials. The two- and five-box trials were eliminated from the analysis because total errors were at floor ($M = 0.26$, $SD = 0.57$) and ceiling ($M = 8.40$, $SD = 4.32$) levels, respectively. A ratio was used because a significant number of children were not administered the four-box trials due to the discontinue rule.

Preschool continuous performance test (P-CPT). The P-CPT (Kerns & McInerney, 2007) was a computerized task in which children were required to respond to the appearance of a target stimulus by pressing a computer button and to refrain from responding to the appearance of a number of different nontarget stimuli. Children were introduced to cartoon pictures of nine farm animals (e.g., dog, pig, sheep, etc.). Each animal was paired with a corresponding animal sound (e.g., the dog was paired with a barking sound). Children were informed that the farmer fed all of the animals except for the target animal (e.g., sheep) and that it was their job to feed the target animal by pressing the space bar on a computer keyboard every time the target appeared. Children also were instructed to avoid pressing the space bar for the nontarget animals because these animals were no longer hungry. The task was preceded by a 30-second practice session in which the animals appeared individually in the middle of the computer screen in 1.5-second spans, and corrective feedback was provided as needed. The task lasted 5 minutes and included 200 stimulus presentations of which 29 were targets. When necessary, the experimenter encouraged children to stay on task. Performance was measured both in terms of working memory and in terms of inhibition. Working memory was measured with the number of omission errors (i.e., not responding to the target animal). Inhibition was measured with: a) the number of commission errors (i.e., responding to a nontarget animal) and b) a ratio of the number of target responses (i.e., hits) to the number of total responses (i.e., hits plus commission errors).

Boy-girl Stroop. The Boy-Girl Stroop (Kerns & McInerney, 2007; adapted from Diamond, Kirkham, & Amso, 2002) was a computerized task designed to measure inhibition. Children were instructed to say "boy" when a picture of a cartoon girl appeared on the screen and to say "girl" when a picture of a cartoon boy appeared on the screen. A short practice session preceded the task in which corrective feedback was provided as needed. The task consisted of 20 pictures, with each type of picture appearing 10 times and never more than 3 times in succession. Immediately following the task, children were asked to explain the rules to confirm that the task had been understood. Two children were unable to correctly explain the rules, and their scores were excluded from the data analysis. Performance was measured with the number of times children correctly labeled each picture, and self-corrections were counted as incorrect.

Tower of Hanoi (ToH). The ToH (Kerns & McInerney, 2007) was a computerized task designed to measure inhibition (Wiebe et al., 2008). Children were instructed to transfer three monkeys of graduated size from a starting position to a specified goal state while observing a number of rules. The procedure for administering the ToH was adapted from Welsh (1991). On the computer screen, children were presented with the three monkeys and three equally spaced trees of the same size. The rightmost tree was classified as the goal tree and had bananas at its base. A short practice session preceded the task in which children were instructed in how to move the monkeys between the trees through the use of a touchscreen monitor. At this time, children also were informed of the following two rules: a) Only one monkey could be moved at a time, and b) smaller monkeys could be placed on top of bigger monkeys, but bigger monkeys could not be placed on top of smaller monkeys. Next, children were informed that the goal of the task was to transfer all of the monkeys to the rightmost tree in the fewest moves possible so that the biggest monkey (i.e., daddy) was on the bottom, the mid-sized monkey (i.e., mommy) was in the middle, and the smallest monkey (i.e., baby) was on top. A small picture of this goal state was displayed at the bottom of the screen throughout the task as a reminder.

The task began with a trial that could be solved in 2 moves and progressively increased in 1-move increments up to a 6-move trial. The maximum number of moves allowed for each trial was 10 moves plus the minimum number of moves required to solve the trial (e.g., a maximum of 13 moves were allowed on the 3-move trial). As long as children solved the trial in the maximum number of moves allowed, it was considered correct. If children were unable to solve a particular trial within the maximum number of moves allowed, the trial ended and was marked as incorrect, and a new trial started. The task was discontinued when children were unable to solve two consecutive trials. Performance was measured with a ratio of the total number of moves for all administered trials to the number of correct trials.[2] The ratio indicated the average number of moves for each correct trial.

Go/No-Go. The Go/No-Go (Kerns & McInerney, 2007) was a computerized task in which children were required to respond to the appearance of

---

[2] Similar results emerged when performance was measured both in terms of number of correct trials and in terms of number of moves for two- and three-move trials combined. Because illegal moves were not recorded, we were unable to test a ratio of the number of illegal moves to the total number of moves (see Wiebe et al., 2008).

target stimuli by pressing a computer button and to refrain from responding to the appearance of nontarget stimuli. The task was divided into four blocks: a baseline block and three test blocks. Each block consisted of 25 stimuli, lasted approximately 45 seconds, and directly followed the previous block. For the baseline block, children were instructed to press the space bar on a computer keyboard every time a picture of a cartoon dog appeared on the screen. They were told that pressing the space bar would scratch the dog's ears and that the dog really liked having his ears scratched. The purpose of the baseline block was to build a prepotent response to touch the space bar for the dog. Before the first test block, children next were introduced to a cartoon koala that also would be appearing on the screen. They were instructed now to only press the space bar for the koala and to not press the space bar for the dog. They were told that pressing the space bar would scratch the koala's tummy and that the dog was tired and needed a rest. Prior to the start of each test block, the experimenter ensured that children could identify both the target and the nontarget stimuli. Corrective feedback was not provided during the test blocks. The second and third test blocks followed the same procedure as the first test block, with children instructed to respond to the dog in the second block and the koala in the third block. For each test block, the dog and koala appeared in pseudorandom order and location, and the target stimulus appeared 13 times while the nontarget stimulus appeared 12 times.

Performance was measured both in terms of inhibition and in terms of set shifting. Inhibition was measured with: a) the number of commission errors (i.e., responding to nontarget stimuli) during Test Block 1, and b) a ratio of the number of target responses (i.e., hits) to the number of total responses (i.e., hits plus commission errors) during Test Block 1. Test Blocks 2 and 3 were not used to measure inhibition due to the possibility that later Go/No-Go blocks make strong demands on attention flexibility (Ozonoff, Strayer, McMahon, & Filloux, 1994). Therefore, Test Blocks 2 and 3 both were used as indicators of set shifting. Performance was measured with a ratio of the number of target responses to the number of total responses.

Border version of the Dimensional Change Card Sort (DCCS). The DCCS was classified as a measure of set shifting (Jacques, Zelazo, Kirkham, & Semcesen, 1999), and the procedure for administering the task followed Zelazo (2006). Children were asked to sort two types of picture cards (e.g., rabbits and boats), both of which could be categorized by two dimensions: color and shape. The experimenter identified both dimensions for children and then randomly chose one dimension (e.g., color) as the sorting rule for the preswitch phase. The preswitch phase began with two practice trials in which corrective feedback was provided as needed. Six preswitch trials were then administered. For each trial, the experimenter repeated the rules of the task, did not provide corrective feedback, and never presented the same type of card on more than two consecutive trials. Immediately following the sixth preswitch trial, the postswitch phase began. In the postswitch phase, the experimenter explained that a new game was going to be played and instructed children to begin sorting the cards by the second dimension (e.g., shape). The postswitch phase consisted of six trials, and the procedure was the same as the preswitch phase. Children who correctly sorted at least five of the six cards in the postswitch phase moved on to the border phase. In the border phase, children were instructed to sort cards with a 6-millimeter black border by one dimension (e.g., color) and cards without a border by the second dimension (e.g., shape). The border phase consisted of six trials, and the procedure was the same as the two previous phases. Performance was measured with the number of correctly sorted cards for all trials administered.

Peabody Picture Vocabulary Test-third edition (PPVT-III). Receptive vocabulary was assessed with the PPVT-III (Dunn & Dunn, 1997). The experimenter stated a word, and children had to point to the corresponding picture out of four choices. The task ended when children made at least 8 errors on a set of 12 words. Performance was measured in terms of raw scores (i.e., ceiling item – total errors).

RESULTS

Data Preparation

Preliminary screening of the P-CPT and Go/No-Go data revealed that a majority of children ($n = 107$ and $n = 74$, respectively) had extremely short response times (i.e., <150 milliseconds) for at least one nontarget stimulus directly following a target stimulus (average of 4.6% of P-CPT responses and an average of 5.6% of Go/No-Go responses). A review of the video recordings led us to believe that these fast responses were the result of children who had initiated a late response to a target stimulus or of children who had held the computer button down longer than necessary, thereby recording a commission error when their performance suggested a correct response. Consequently, these “fast-response” commission errors were excluded from children’s P-CPT and Go/No-Go scores because they did not accurately represent children’s performance on these tasks.

All variables were screened for univariate and multivariate outliers, for skewness, and for kurtosis using the software package Predictive Analytics SoftWare Statistics 18.0. Five outliers were found for P-CPT commission errors, and three outliers were found for the Boy-Girl Stroop. These outlier values were replaced with the highest remaining score ±1 under the assumption that children's true scores were extreme on these tasks (Tabachnick & Fidell, 2007, p. 77). A review of the video recordings confirmed that these children were engaged but displayed great difficulty with the particular task. Mahalanobis distance did not reveal any multivariate outliers. P-CPT commission errors were positively skewed and leptokurtic (i.e., values ≥ 2). Attempts to correct for this departure from normality (e.g., logarithmic transformation) yielded results similar to analyses with the untransformed data. Therefore, for ease of interpretation, analyses were conducted with the untransformed data. All other variables were reasonably distributed with only minor departures from normality.

Descriptive Statistics

Table 1 displays descriptive statistics for all manifest variables. Girls performed significantly better than boys as indexed by P-CPT commission

TABLE 1
Descriptive Statistics for Executive Function Measures, Receptive Vocabulary, Age, and Maternal Education

Variable	*N*	*M*	*SD*	*Range*	*Skew. (SE)*	*Kurt. (SE)*
Backward Digit Span highest level	128	0.81	1.14	0–4	0.97 (0.21)	–0.55 (0.43)
Backward Word Span highest level	125	1.14	1.25	0–4	0.59 (0.22)	–0.92 (0.43)
Boxes hits/total responses	120	0.73	0.13	0.50–1.00	0.56 (0.22)	–0.37 (0.44)
P-CPT omission errors	119	7.80	6.39	0–24	0.70 (0.22)	–0.54 (0.44)
P-CPT commission errors	119	6.24	7.67	0–31	2.17 (0.22)	4.10 (0.44)
P-CPT hits/total responses	119	0.79	0.18	0.28–1.00	–1.20 (0.22)	0.59 (0.44)
Boy-Girl Stroop correct trials	118	15.86	3.46	5–20	–1.29 (0.22)	1.52 (0.44)
Tower of Hanoi moves/correct trials	125	0.07	0.04	0.00–0.19	0.35 (0.22)	–0.40 (0.43)
GNG Block 1 commission errors	110	2.39	2.03	0–8	0.78 (0.23)	–0.27 (0.46)
GNG Block 1 hits/total responses	110	0.75	0.22	0.00–1.00	–1.51 (0.23)	3.17 (0.46)
GNG Block 2 hits/total responses	110	0.74	0.20	0.00–1.00	–1.00 (0.23)	1.84 (0.46)
GNG Block 3 hits/total responses	110	0.79	0.20	0.09–1.00	–1.44 (0.23)	2.31 (0.46)
DCCS correct trials	125	12.15	4.15	6–18	–0.62 (0.22)	–1.33 (0.43)
PPVT-III raw score	127	67.87	18.24	25–102	–0.23 (0.22)	–0.69 (0.43)
Age in months	129	50.45	6.97	36.0–68.2	0.40 (0.21)	–0.45 (0.42)
Maternal education	129	3.61	0.87	2–5	0.19 (0.21)	–0.79 (0.42)

P-CPT = Preschool Continuous Performance Test; GNG = Go/No-Go; DCCS = Dimensional Change Card Sort; PPVT-III = Peabody Picture Vocabulary Test-Third edition.

errors, P-CPT hit ratio, Go/No-Go Block 1 commission errors, Go/No-Go Block 1 hit ratio, and Go/No-Go Block 3 hit ratio. All other variables had no significant sex differences. Variations in task sample size were due to children's request to end a task or to their failure to understand or to perform a task. In total, 6% of the data were missing and handled with full information maximum likelihood estimation in Amos 18.0 (Arbuckle, 2009). Zero-order correlations between all manifest variables are presented in Table 2. For ease of interpretation, variables that were scored in terms of error responses (i.e., P-CPT omission errors, P-CPT commission errors, and Go/No-Go Block 1 commission errors) were reverse-scored to correspond to the other variables that were scored in terms of correct responses. In most cases, stronger correlations were found among variables that shared a common EF process (i.e., working memory, inhibition, or set shifting). All variables were significantly correlated with receptive vocabulary. With the exceptions of the Boxes task and the Boy-Girl Stroop, all variables also were significantly correlated with age, with older children performing better than younger children. The DCCS was the only variable significantly correlated with maternal education.

Confirmatory Factor Analyses

In keeping with our goals of replicating and extending the analysis conducted by Wiebe et al. (2008), statistical analyses were separated into two distinct CFAs. The first CFA (see Figure 1) examined the unitary EF factor structure with performance indicators similar to those used by Wiebe et al. (2008). The second CFA estimated and compared four EF models (see Figure 2) in which performance indicators were selected to provide a clear distinction among working memory, set shifting, and inhibition factors. Based on the covariance matrix of the manifest EF variables, Amos 18.0 (Arbuckle, 2009) was used to fit all models with full information maximum likelihood estimation. Estimated means and intercepts were selected to account for missing data. To scale each model for estimation, a single manifest variable was fixed to load at 1.00 for each latent factor. For all estimated models, the error variances between the Backward Span tasks were correlated to account for similarities between the two tasks in scaling and in method variance. All other error variances were uncorrelated.[3]

---

[3] Scaling and method variance also were similar for Go/No-Go Block 2 hit ratio and Go/No-Go Block 3 hit ratio. However, correlating the error variances between these two variables resulted in nonsignificant values at $p > .05$. Therefore, the variance values for these two Go/No-Go blocks remained unconstrained for all estimated models.

TABLE 2
Zero-Order Correlations Among Executive Function Measures, Receptive Vocabulary, Age, and Maternal Education ($N = 129$)

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Backward Digit Span	—	—	—	—	—	—	—	—	—	—	—	—	—
2. Backward Word Span	.78**	—	—	—	—	—	—	—	—	—	—	—	—
3. Boxes hit ratio	.23*	.24*	—	—	—	—	—	—	—	—	—	—	—
4. P-CPT omissions	.38**	.50**	.19*	—	—	—	—	—	—	—	—	—	—
5. P-CPT commissions	.20*	.26**	.09	.28**	—	—	—	—	—	—	—	—	—
6. P-CPT hit ratio	.29**	.35**	.13	.58**	.90**	—	—	—	—	—	—	—	—
7. Boy-Girl Stroop	12	.26**	.19*	.15	.24*	.24*	—	—	—	—	—	—	—
8. Tower of Hanoi	.18*	.18†	.15	.20*	.30**	.35**	.09	—	—	—	—	—	—
9. GNG1 commissions	.13	.15	.15	.24*	.39**	.42**	.31**	.17†	—	—	—	—	—
10. GNG1 hit ratio	.23*	.31**	.15	.37**	.23*	.33**	.20*	.19†	.43**	—	—	—	—
11. GNG2 hit ratio	.30**	.41**	.12	.48**	.34**	.47**	.28**	.17†	.18†	.22*	—	—	—
12. GNG3 hit ratio	.18†	.22*	.10	.29**	.17†	.27**	.16†	.28**	.14	.27**	.25**	—	—
13. DCCS	.45**	.59**	.19*	.44**	.09	.22*	.19*	.13	.19†	.40**	.40**	.26**	—
PPVT-III raw score	.46**	.48**	.18*	.29**	.35**	.39**	.22*	.29**	.30**	.36**	.33**	.26**	.53**
Age in months	.48**	.52**	.09	.45**	.18*	.31**	.12	.26**	.22*	.30**	.47**	.28**	.52**
Maternal education	–.03	.06	.03	–.06	.05	–.02	.01	–.03	.01	.05	–.13	.04	.23**

Note. Variables 4, 5, and 9 are reverse-scored to ease interpretation.

P-CPT = Preschool Continuous Performance Test; GNG = Go/No-Go with block number; DCCS = Dimensional Change Card Sort; PPVT-III = Peabody Picture Vocabulary Test-Third edition.

$^{\dagger}p \leq .10$. $^{*}p \leq .05$. $^{**}p \leq .01$.

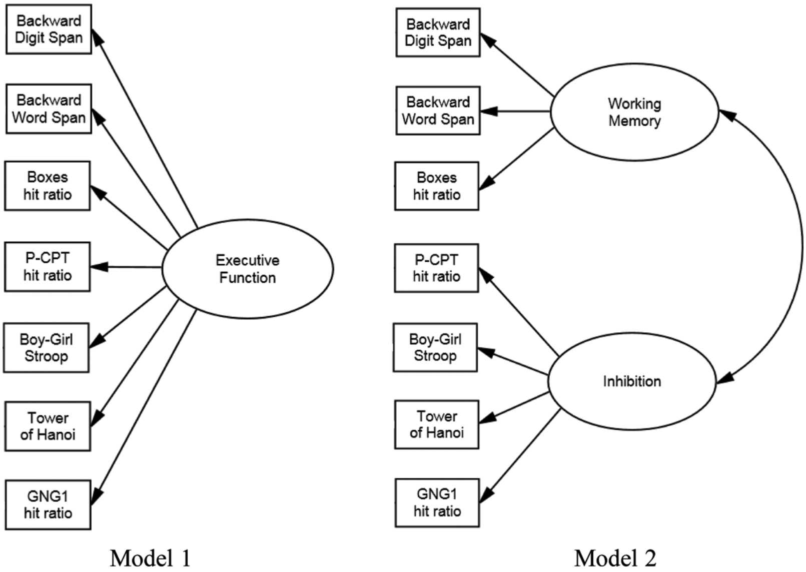

FIGURE 1 First model series. P-CPT = Preschool Continuous Performance Test; GNG1 = Go/No-Go Block 1.

Model fit was evaluated using the chi-square goodness-of-fit test (Loehlin, 1998), chi-square/degrees of freedom (Bollen, 1989), the comparative fit index (CFI; Bentler, 1990), and the root mean square error of approximation (RMSEA; Steiger, 1990). Good model fit was associated with a low chi-square value, chi-square/degrees of freedom ≤ 3, a CFI $\geq .95$, and an RMSEA $\leq .06$ (Kline, 2005). Nested model comparisons were evaluated using the chi-square difference test. If two nested models fit the observed data equally well, the simpler model was preferred when comparison to a more complex model did not differ significantly at $p \leq .05$ (Bollen, 1989). Model comparisons also were evaluated using Akaike's information criterion (AIC; Bozdogan, 2000), which penalizes more complex models by accounting for the number of estimated model parameters. Lower AIC values represented better model fit (Kline, 2005).

Replication of the unitary factor structure. The first model series (see Figure 1) sought to replicate the unitary EF factor structure that Wiebe et al. (2008) found in their preschool sample. To draw the best comparison between studies, tasks and performance indicators were matched as closely as possible to those used by Wiebe et al. (2008). To this end, the Backward

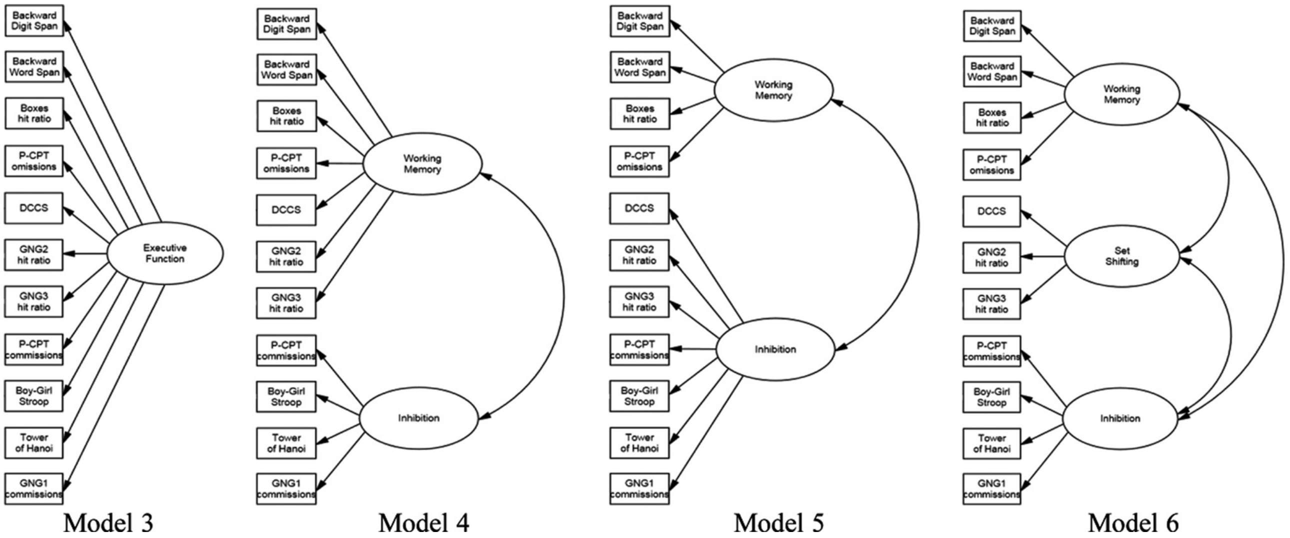

FIGURE 2 Second model series. P-CPT = Preschool Continuous Performance Test; DCCS = Dimensional Change Card Sort; GNG = Go/No-Go with block number.

Digit Span, the Boxes task, the P-CPT, and the ToH were selected to correspond to similar tasks used by Wiebe et al. (2008). In addition, substitutions were made for tasks with no direct match. Specifically, the Boy-Girl Stroop was substituted for the Shape School task (Espy, 1997) used by Wiebe et al. (2008) on the grounds that both tasks required children to inhibit a prepotent verbal response, and the Go/No-Go was substituted for the Visual Attention task (Korkman, Kirk, & Kemp, 1998) on the grounds that both tasks required children to simultaneously respond to target stimuli and ignore distracter stimuli. Tasks used by Wiebe et al. (2008) for which there were no suitable matches or substitutions included the Delayed Alternation task, the Delayed Response task, and the Statue task. Finally, the Backward Word Span was included as an additional working memory indicator. All matched tasks, task substitutions, and task additions are summarized in Table 3.

The specific EF models tested in the replication series included a single-factor model in which all indicators loaded on a unitary EF factor (Model 1) and a two-factor model with working memory and inhibition factors (Model 2; see Figure 1). The overall fit between both models and the observed data was good (see Table 4). In addition, the chi-square difference between Models 1 and 2 indicated that both models fit the observed data equally well. Model 1, however, had a lower AIC value and was more parsimonious than Model 2. As a result, Model 1 was preferred over Model 2, supporting the unitary EF factor structure found by Wiebe et al. (2008). Estimates for Model 1 are displayed in Table 5; there was no indication of problems with residual values (e.g., large or negative values), and the error variance correlation between the Backward Span tasks was significant ($r = 0.71$, $p < .01$).

Testing a three-factor structure. The second model series (Figure 2) was created to examine whether the EF factor structure in preschoolers a) is affected by the selection of performance indicators, and b) is similar to the three-factor structure that has been found both in school-aged children (Lehto et al., 2003) and in adults (Miyake et al., 2000). In contrast to the first model series, performance indicators in the second model series were selected to predominantly capture active rule maintenance for the working memory factor and the suppression of prepotent responses for the inhibition factor. To this end, the P-CPT omission errors indicator was included as a measure of working memory, and commission errors replaced hit ratios as inhibition indicators for both the P-CPT and the Go/No-Go Block 1. In addition, the second model series included three performance indicators designed to measure set shifting: DCCS, Go/No-Go Block 2 hit ratio, and Go/No-Go Block 3 hit ratio. Performance indicators for the Backward

TABLE 3
Comparison of Tasks and Performance Indicators Used in Wiebe et al. (2008), the First Model Series, and the Second Model Series

Wiebe et al. (2008)	*First model series*	*Second model series*
Six Boxes task: correct/total responses	Boxes task: correct/total responses	Boxes task: correct/total responses
Child CPT: correct/total responses	Preschool CPT: correct/total responses	Preschool CPT: commission errors
Digit Span: maximum length	Backward Digit Span: maximum length	Backward Digit Span: maximum length
Shape School: latency in seconds	Boy-Girl Stroop: correct responses	Boy-Girl Stroop: correct responses
Tower of Hanoi: illegal/total moves	Tower of Hanoi: total moves/correct trials	Tower of Hanoi: total moves/correct trials
Visual Attention: correct responses – errors	Go/No-Go Block 1: correct/total responses	Go/No-Go Block 1: commission errors
Delayed Alternation: correct searches	—	—
Delayed Response: correct searches	—	—
Statue: 5-second epochs	—	—
—	Backward Word Span: maximum length	Backward Word Span: maximum length
—	—	Preschool CPT: omission errors
—	—	DCCS: correct responses
—	—	Go/No-Go Block 2: correct/total responses
—	—	Go/No-Go Block 3: correct/total responses

CPT = Continuous Performance Test; DCCS = Dimensional Change Card Sort.

Span tasks, the Boxes task, the Boy-Girl Stroop, and the ToH remained unchanged from the first model series (see Table 3).

The specific EF models tested in the second model series included a single-factor model in which all indicators loaded on a unitary EF factor (Model 3), a pair of two-factor models with working memory and inhibition factors in which the set shifting indicators loaded either on the working memory factor (Model 4) or on the inhibition factor (Model 5), and a three-factor model with working memory, set shifting, and inhibition factors (Model 6; see Figure 2). The model fit indices (see Table 4) indicated that each model provided a good or an adequate fit to the observed data. However, the covariance matrix for Model 6 was not positive definite, which meant that the factor solution for Model 6 was inadmissible (Arbuckle, 2009); this was most likely due to an inflated correlation between the

TABLE 4
Model Fit Indices for Confirmatory Factor Analysis Models ($N = 129$)

Model (number of factors)	*df*	χ^{2a}	χ^2/df^b	CFI^c	$RMSEA^d$	AIC^e	*Model comparison*	Δdf	$\Delta\chi^{2f}$
First model series									
1. Unitary executive function (1)	13	9.55	0.74	1.00	.000	53.55	—	—	—
2. Working memory & inhibition (2)	12	8.61	0.72	1.00	.000	54.61	*Model 1* vs. Model 2	1	0.94
Second model series									
3. Unitary executive function (1)	43	61.98^*	1.44	0.94	.059	129.98	—	—	—
4. High working memory loading (2)	42	43.41	1.03	1.00	.016	113.41	Model 3 vs. *Model 4*	1	18.57^{**}
5. High inhibition loading (2)	42	61.93^*	1.47	0.93	.061	131.93	*Model 4* vs. Model 5	0	18.52^{**}
6. Working memory, set shifting, & inhibition (3)[g]	40	43.24	1.08	0.99	.025	117.24	*Model 4* vs. Model 6	2	0.17

Note. The preferred models are italicized.

[a] Lower values indicated better model fit; values with $p \leq .05$ indicated that the model did not fit the data better than a saturated model.

[b] Values ≤ 3 indicated good model fit.

[c] Values ≥ 0.95 indicated good model fit; values ≥ 0.90 indicated adequate model fit.

[d] Values $\leq .06$ indicated good model fit; values $\leq .08$ indicated adequate model fit.

[e] When comparing models, lower values indicated the better model fit.

[f] Values with $p \leq .05$ indicated that the simpler model was significantly less satisfactory than the comparatively complex model.

[g] Not positive definite covariance matrix.

$^*p \leq .05$. $^{**}p \leq .01$.

working memory and set shifting factors ($r = 1.03$, $p < .001$). Therefore, excluding Model 6, model comparisons were conducted to determine the best-fitting model. The chi-square difference between Models 3 and 4 indicated that Model 4 fit the observed data significantly better than Model 3, and Model 4 also had a lower AIC value than Model 3. As a result, Model 4 was preferred over Model 3. Next, Model 4 was compared to Model 5. Because both two-factor models had the same degrees of freedom, a chi-square difference test could not be conducted. Nevertheless, due to a lower AIC value and better model fit indices, Model 4 remained the preferred model over Model 5. Finally, although Model 6 was inadmissible, when compared with Model 4, Model 4 was still preferred in terms of parsimony and a lower AIC value. As a result, Model 4 was the overall preferred model, contrasting the results from the first model series. Estimates for Model 4 are displayed in Table 5; the correlation between the working memory and inhibition factors was significant ($r = 0.56$, $p < .001$). In addition, there was no indication of problems with residual values, and

TABLE 5
Summary of Results for the Best Fitting Models for the First (Model 1) and Second (Model 4) Model Series ($N = 129$)

Manifest variable	*Latent factor*	*B*	*SE*	*Critical ratio*	*ß*	R^2
Model 1						
Backward Digit Span	Executive Function	1.00	—	—	.45	.20
Backward Word Span	Executive Function	1.37***	0.22	6.14	.56	.31
Boxes hit ratio	Executive Function	0.09*	0.04	2.43	.33	.11
P-CPT hit ratio	Executive Function	0.23***	0.07	3.36	.64	.41
Boy-Girl Stroop	Executive Function	2.86**	1.01	2.83	.42	.17
Tower of Hanoi	Executive Function	0.04**	0.01	2.90	.43	.18
GNG1 hit ratio	Executive Function	0.23**	0.07	3.15	.53	.28
Model 4						
Backward Digit Span	Working Memory	1.00	—	—	.58	.34
Backward Word Span	Working Memory	1.42***	0.16	8.85	.75	.57
Boxes hit ratio	Working Memory	0.06**	0.02	2.70	.29	.09
P-CPT omissions	Working Memory	6.42***	1.26	5.10	.67	.44
DCCS	Working Memory	4.39***	0.83	5.26	.70	.49
GNG2 hit ratio	Working Memory	0.18***	0.04	4.75	.61	.37
GNG3 hit ratio	Working Memory	0.12***	0.04	3.30	.38	.14
P-CPT commissions	Inhibition	1.00	—	—	.65	.42
Boy-Girl Stroop	Inhibition	0.34***	0.09	3.57	.48	.23
Tower of Hanoi	Inhibition	0.00***	0.00	3.27	.41	.17
GNG1 commissions	Inhibition	0.25***	0.06	3.96	.60	.36

P-CPT = Preschool Continuous Performance Test; GNG = Go/No-Go with block number; DCCS = Dimensional Change Card Sort.

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

the error variance correlation between the Backward Span tasks was significant ($r = 0.64$, $p < .001$).

DISCUSSION

The goals of the present study were to: a) replicate the unitary factor structure found by previous CFA studies of EF in preschool children, b) examine the diversity of the EF structure in preschool children by including additional working memory and inhibition performance indicators, and c) explore whether set shifting constitutes a latent factor of EF in preschool children. We discuss the pertinent findings in turn.

Replication of the Unitary Factor Structure

The first model series of the present study sought to replicate the unitary EF factor structure found by Wiebe and colleagues (2008) by including similar performance indicators of working memory and inhibition. The results of the CFA indicated that a single-factor model of EF was preferred over a two-factor model consisting of working memory and inhibition factors. This finding not only replicated the findings by Wiebe et al. (2008), but it also was consistent with other CFA studies of EF with preschool children (Hughes et al., 2010; Wiebe et al., 2011).

Wiebe et al. (2008) suggested that EF might constitute a unitary construct in preschoolers because prepotent responses are overcome not by inhibition, but by enhancing the activation of correct stimulus–response relations. According to Wiebe et al. (2008), working memory plays a central role in preschoolers' EF performance, and inhibition does not emerge as an independent factor. Nevertheless, in a more recent study, Wiebe et al. (2011) have not excluded the possibility of a two-factor EF structure in preschool children but instead have stated that "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" (p. 448). In fact, in the current model series and in both CFA studies by Wiebe et al. (2008, 2011), the two-factor EF model that specified independent working memory and inhibition factors fit the data equally well as the single-factor model. We also note that, due to model constraints, Hughes et al. (2010) were unable to compare their single-factor EF model to other more complex models. As a consequence, previous CFA findings with preschoolers do not rule out the possibility that EF in preschool children extends beyond a unitary structure.

We believe that the unitary EF factor structure supported by previous CFA findings with preschool children has, in large part, been the result of the particular tasks and performance indicators that were selected to represent the working memory and inhibition factors. Specifically, in all previous CFA findings in which multiple EF models were compared (present study; Wiebe et al., 2008, 2011), the distinction between the working memory and inhibition factors might have been blurred due to an overlap in working memory and inhibition task demands. We tested this possibility by examining whether a more diverse EF structure would emerge when performance indicators were selected that would, at least in theory, establish a greater distinction between working memory and inhibition demands.

Diversity of EF in Preschoolers

The selection of performance indicators in the second model series was guided by the goal of separating active rule maintenance and the inhibition of prepotent responses. Specifically, the complexity of the model structure was varied to examine whether the organization of EF in preschool children was in line with the structure observed in school-aged children and adults. The results of the CFA indicated that a two-factor model consisting of working memory and inhibition processes best represented the structure of EF in preschool children. Whereas these findings supported a diversified EF structure similar to research with school-aged children (Huizinga et al., 2006; Lehto et al., 2003; van der Sluis et al., 2007) and adults (Miyake et al., 2000), they were in contrast both to the first model series of the present study and to previous CFA studies that found that EF was best represented as a unitary construct in preschoolers (Hughes et al., 2010; Wiebe et al., 2008, 2011). Moreover, in previous CFA findings with preschoolers, the favored single-factor models always fit the data equally as well as the two-factor models with working memory and inhibition factors. Our findings are strengthened by the fact that our preferred two-factor model of EF outperformed the single-factor model in terms of both model fit and model comparison values.

Compared with the findings in the first model series, the findings in the second model series highlight the theoretical importance of performance indicator selection in CFA studies of EF in preschool children. On the one hand, when we selected performance indicators of working memory and inhibition similar to those chosen by Wiebe et al. (2008), we were able to replicate the favored unitary EF factor structure in preschool children. On the other hand, when we selected performance indicators on the basis of creating a more clear-cut distinction between working memory and inhibition processes, a differentiated EF structure was preferred over a unitary structure. The overall message from these findings seems to be that the structure of EF that emerges from CFA is influenced by task and performance indicator selection. A practical implication of these findings is that it is an empirical question whether the single- or two-factor EF model in preschoolers is more useful in the prediction of behavior problems, social skills, and academic school readiness (Wiebe et al., 2011).

The significant correlation between the working memory and inhibition factors in our preferred two-factor model of EF indicates that the two EF processes are separable but related components of EF in preschool children. This distinction is consistent with the interactive framework of EF proposed by Roberts and Pennington (1996) in which performance on EF tasks depends on the dynamic interaction between working memory and inhibition processes. Acting like a pulley system, the interactive framework of EF allows preschoolers to engage one process (e.g., working memory) more than the other process (e.g., inhibition) depending on the ongoing demands of a particular situation. For example, on the P-CPT, children are likely to rely more on working memory to refrain from making an omission error, but then are likely to rely more on inhibitory control to refrain from making a commission error. Moreover, the interactive framework stipulates that performance on inhibition tasks depends on the strength of the prepotent response. It is likely that our selection of a relatively homogenous set of prepotent response inhibition tasks made it possible to distinguish a differentiated EF structure in preschoolers. For instance, as found in the preferred two-factor model, when commission errors for both the P-CPT and the Go/No-Go Block 1 were used as inhibition indicators, the correlation between the working memory and inhibition factors was .56. However, replacing the commission error indicators with hit ratio indicators similar to those used by Wiebe et al. (2008) results in an increased factor correlation of .86, indicating more commonality and less separation between EF factors. Similar results are found for the other two-factor models that were tested in the present study. Taken together, these findings suggest that the use of commission errors as indicators of prepotent response inhibition leads to a separation between EF processes in preschoolers.

Latent Set Shifting Factor

Whereas previous CFA studies of EF in school-aged children (Lehto et al., 2003) and adults (Miyake et al., 2000) have supported a tripartite structure of EF that included working memory, inhibition, and set shifting factors, the results of our CFA did not support a set shifting factor in preschool children. Although the three-factor EF model provided a good fit to the data, the covariance matrix of the model was not positive definite. As such, the model solution was inadmissible and, therefore, could not meaningfully be interpreted (Arbuckle, 2009).

These findings are likely due to an overlap in shared variance between the working memory and set shifting tasks, as evident in the inflated correlation between the working memory and set shifting factors in the three-factor model. Rather than the working memory indicators containing properties of set shifting, it seems more likely that the set shifting indicators contained properties of working memory. For one, in the preferred model (Model 4), the set shifting indicators loaded significantly on the working memory factor. Second, the set shifting indicators used in the present study arguably made strong working memory demands. For instance, the border version of the DCCS makes high demands on active rule maintenance and the operational component of working memory (i.e., recoding of stimulus features; Baddeley, 1996). Previous research also has shown that performance on the standard version of the DCCS is significantly correlated with performance on working memory tasks (Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Zelazo, Müller, Frye, & Marcovitch, 2003, Exp. 3). Further, the later blocks of the Go/No-Go also might have required high working memory demands to keep in mind what animal was the target stimulus and what animal was the distracter stimulus. Future research is needed to examine these claims.

Conclusion

By using CFA to examine the composition of EF, this study demonstrated that working memory and inhibition are distinguishable as latent variables in preschool children. These findings have important implications for the conceptualization of the development of EF. On the one hand, these findings are consistent with the proposal that working memory and inhibition are separable components of EF in young children (Diamond, 2002). However, working memory and inhibition were significantly correlated, thus indicating both unity and diversity of EF (Miyake et al., 2000). On the other hand, our findings from the second model series with more clearly separated working memory and inhibition indicators are inconsistent with the proposal that EF in preschool children constitutes a unitary system (Hughes et al., 2010; Wiebe et al., 2008, 2011), and they are at odds with the suggestion that EF in preschool children lacks coherence (Garon et al., 2008). The conceptualization of EF development has important implications for understanding a variety of developmental disorders, which have been characterized by impairments in different components of EF (Pennington & Ozonoff, 1996).

Some limitations of the current analysis should be taken into consideration. First, in relation to previous CFA studies of EF in typically developing preschoolers (Hughes et al., 2010; Wiebe et al., 2008, 2011), the sample of the present study was high in parental education; although this may have influenced the results of the CFA, we were still able to replicate findings in the first model series that were consistent with previous research. Another issue concerns the age range of children in the present study. Whereas our preschool sample included children between 3 and 5 years of age, Wiebe et al. (2008) used a sample of children between 2 and 6 years of age, and Wiebe et al. (2011) used a sample of 3-year-old children. Although Wiebe et al. (2008) did not specify the number of children tested at each age, if younger children dominated their sample, then the results of both studies by Wiebe et al. (2008, 2011) may be more representative of younger preschoolers. In fact, invariance testing by Wiebe et al. (2008) revealed that their unitary EF factor accounted for more variance in younger preschoolers compared with older preschoolers. This may indicate that the structure of EF differentiates during the early preschool years. The results of the present study support the notion of EF as an emerging ability in the preschool years and suggest that tasks with stronger distinctions between working memory and inhibition demands are able to detect this emerging EF structure. Unfortunately, due to sample size constraints, we were unable to adequately conduct invariance testing for different age groups of children.[4] Future research with large samples of younger and older preschoolers is needed to investigate this possibility.

A third issue concerns the low parameter estimate and squared multiple correlation values for the Boxes task. Interestingly, Wiebe et al. (2011) found comparably low values in their CFA model solutions for a similar task, the Nine Boxes task (adapted from Diamond, Prevor, Callender, & Druin, 1997). One reason for these low values may be that, in addition to working memory capacity, self-ordered search tasks tap into independent processes, such as proactive interference (Friedman & Miyake, 2004) or substitution (Ecker, Lewandowsky, Oberauer, & Chee, 2010). This possibility could be explored in future studies by including additional measures of proactive interference and updating and then examining the loading patterns of the different working memory tasks.

There also are methodological implications for future studies of EF in preschool children. Even though CFA is an important tool to address the

---

[4] For exploratory purposes, we conducted a test of invariance comparing younger (< 4 years old; $n = 55$, $M_{age} = 3;8$, $SD_{age} = 3$ months) and older (> 4 years old; $n = 74$, $M_{age} = 4;7$, $SD_{age} = 5$ months) preschoolers. This test indicated strict measurement invariance, $\chi^2(84, N = 129) = 87.72$, $p > .05$. In addition, a test of invariance for preschoolers divided into boys ($n = 78$, $M_{age} = 4;3$, $SD_{age} = 7$ months) and girls ($n = 51$, $M_{age} = 4;2$s, $SD_{age} = 7$ months) indicated strict measurement invariance for sex, $\chi^2(84, N = 129) = 96.60$, $p > .05$.

task impurity problem, the factor structure of EF detected by this method depends on the type of EF tasks and performance indicators that are entered into the analysis. Differences in task selection and interpretation inevitably result in different factor structures and complicate the comparison of CFA solutions across EF studies (Lehto et al., 2003). The clarification of the structure of EF depends, in addition to CFA, on the use of prudently chosen control tasks (e.g., van der Sluis et al., 2007), experimental manipulations (Zelazo et al., 2003), and theoretical advances that incorporate recent findings from developmental neuropsychology (Garon et al., 2008). The integration of these lines of research in conjunction with the use of different developmental designs (Flynn, O'Malley, & Wood, 2004; Hughes et al., 2010) will then also shed light on the important questions of what the structure of EF reflects and how the organization of EF is accomplished.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Human Early Learning Partnership to Ulrich Müller and by a grant from Research in Early Education and Childhood Health to Kimberly A. Kerns.

REFERENCES

Arbuckle, J. L. (2009). *Analysis of Moment Structures (Amos Version 18.0)* [Computer software]. Crawfordville, FL: Amos Development Corporation.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology (Special Issue: Working Memory)*, *49A*, 5–28.

Bentler, P. M. (1990). Fit indices, LaGrange multipliers, constraint changes, and incomplete data in structural models. *Multivariate Behavioral Research*, *25*, 163–172.

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

Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, *44*, 62–91.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*, 571–593.

Bryant, F. B., & Yarnold, P. R. (1994). Principle-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: American Psychological Association.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *29*, 595–616.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122.

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

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92.

Davis, H. L., & Pratt, C. (1995). The development of theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*, 25–31.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal-lobe function* (pp. 466–503). London, UK: Oxford University Press.

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). Oxford, UK: Oxford University Press.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*, 352–362.

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*, 1–205.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Services.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 170–189.

Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology*, *13*, 495–499.

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 format tasks. *Brain and Cognition*, *41*, 178–199.

Fernyhough, C. (2010). Vygotsky, Luria, and the social brain. In B. W. Sokol, U. Müller, J. I. M. Carpendale, A. R. Young & G. Iarocci (Eds.), *Self and social regulation* (pp. 56–79). New York, NY: Oxford University Press.

Flynn, E. (2007). The role of inhibitory control in false-belief understanding. *Infant and Child Development*, *16*, 53–69.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false-belief understanding and inhibition skills. *Developmental Science*, *7*, 103–115.

Friedman, N. P., & Miyake, A. (2004). The relations between inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Fuster, J. (1989). *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe* (2nd ed.). New York, NY: Raven.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60.

Goldman, P. S., Rosvold, H. E., & Mishkin, M. (1970). Evidence for behavioral impairment following prefrontal lobectomy in the infant monkey. *Journal of Comparative and Physiological Psychology*, *70*, 454–463.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*, 532–560.

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*, 617–644.

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., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*, 1447–1459.

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.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health*, *7*, 131–142.

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.

Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology*, *35*, 770–780.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychological Review*, *17*, 213–233.

Kerns, K. A., & McInerney, R. (2007). *Preschool Tasks* [Computer software]. Victoria, BC, Canada: University of Victoria.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

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.

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H... Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal-lobe functioning. *Developmental Neuropsychology*, *7*, 377–395.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Erlbaum.

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.

Müller, U., Liebermann, D., Frye, D., & Zelazo, P. D. (2008). Executive function, school readiness, and school achievement. In S. K. Thurman & C. A. Fiorello (Eds.), *Applied cognitive research in K–3 classrooms* (pp. 41–83). New York, NY: Routledge.

O'Hearn, K., Asato, M., Ordaz, S., & Luna, B. (2008). Neurodevelopment and executive function in autism. *Development and Psychopathology*, *20*, 1103–1132.

Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism: An information-processing approach. *Journal of Child Psychology and Psychiatry*, *35*, 1015–1031.

Pennington, B. F. (1997). Dimensions of executive functions in normal and abnormal development. In N. A. Krasnegor, G. R. Lyon & P. S. Goldman-Rakic (Eds.), *Development of the prefrontal cortex: Evolution, neurobiology, and behavior* (pp. 265–281). Baltimore, MD: Brookes.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*, 51–87.

Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, *12*, 105–126.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*, 74–81.

Schoemaker, K., Bunte, T., Wiebe, S. A., Espy, K. A., Deković, M., & Matthys, W. (2012). Executive function deficits in preschool children with ADHD and DBD. *Journal of Child Psychology and Psychiatry*, *53*, 111–119.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the investigation of language and thought* (pp. 277–311). Cambridge, MA: MIT Press.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*, 427–449.

Vygotsky, L. S., & Luria, A. R. (1994). Tool and symbol in child development. In J. Valsiner & R. van der Veer (Eds.), *The Vygotsky reader* (pp. 99–175). Oxford, UK: Blackwell.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*, 45–53.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development*, *6*, 59–76.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*, 131–149.

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., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*, 436–452.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*, 1336–1346.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297–301.

Zelazo, P. D., & Müller, U. (2010). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (2nd ed., pp. 574–603). Malden, MA: Blackwell.

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, *68*(Serial No. 264), 11–119.