November 2001

New procedures to assess executive functions in preschool children

K. A. Espy
University of Nebraska-Lincoln, kespy2@unl.edu

P. M. Kaufmann

M. L. Glisky

M. D. McDiarmid

The Clinical Neurospsychologist
2001, Vol. 15, No. 1, pp. 46–58

1385-4046/01/1501-046$16.00

New Procedures to Assess Executive Functions in Preschool Children*

Kimberly Andrews Espy[1], Paul M. Kaufmann[2], Martha L. Glisky[3], and Melanie D. McDiarmid[1]
[1]Southern Illinois University, School of Medicine, Carbondale, IL, USA, [2]Clyde L. Choate Mental Health and Developmental Center, Carbondale, IL, USA, and
[3]Northwestern University Medical School, Chicago, IL, USA

ABSTRACT

Executive functions are difficult to assess in preschool children, yet the preschool period is particularly important, both in the development of behavioral control and of the brain, particularly the prefrontal cortex. Several tasks were adapted from developmental and neuroscience literature and then administered to 98 preschool children (30-, 36-, 42-, 48- and 60-month age groups). Executive function task performance was related largely to age group, but not to sex or intelligence. These tasks, then, were sensitive in this age range and may be useful to delineate distinct cognitive profiles among preschool children with various neurological and developmental disorders.

The assessment of executive functions in young children is controversial. Historically, many researchers considered executive functions to be ‘‘absent’’ in children under 12 years of age (e.g., Smith, 1983). After Chelune and Baer (1986) demonstrated age-related change in performance on cardinal executive function tasks, such as the Wisconsin Card Sorting Test (Heaton, 1981), the emergence of executive skill was ‘‘moved down’’ to early school age. Such pronouncements fit with the popular view that young children lack inhibitory control, are distractible, and have difficulty shifting among tasks. Concurrently, developmental psychologists, who have noted the relation of performance on Piagetian tasks, such as A-not-B, and the function of the prefrontal cortex in humans and animals (Diamond, 1990), have suggested that executive skills originate in infancy (Diamond, 199l). Similar to other cognitive skills such as language, executive functions, although not present in their fully developed form, can be measured across the early life span, if developmentally appropriate tasks are used that take into account the more limited behavioral repertoire of infants and young children.

Developing executive function tasks for use in preschool children is important for two reasons. First, many common disorders manifest prior to school age (e.g., Attention Deficit Hyperactivity Disorder, genetic abnormalities, prematurity, toxic exposures). In school age children, results from recent studies indicate that executive function measures are sensitive to prefrontal damage in diverse clinical populations, such as closed-

*This research was supported, in part, by grants to the first author from the University of Arizona Foundation (I/5ESPY) and the Southern Illinois University Office of Research Development Administration, Special Research Program. The authors thank the staff from the Beginning School, Presbyterian Child Development Center, and the Child Development Laboratory and the participating families.
Address correspondence to: Kimberly Andrews Espy, Department of Psychiatry, Mailcode 6503, Southern Illinois University School of Medicine, Carbondale, IL, 62901-6503, USA. E-mail: kespy@siumed.edu
Accepted for publication: October 30, 2000.

head injury (Levin et al., 1994) and fetal alcohol syndrome (Mattson, Goodman, Caine, Delis, & Riley, 1999). Few measures exist to assess executive skills in children under age 6 years, and even fewer for children under 3 years of age. Welsh, Pennington, Rouse, Ozonoff, and McCabe (1990) provided some of the first examples in the pediatric neuropsychology literature of adapting developmental tasks (e.g., Tower of Hanoi, Visual Search) for use in investigations with preschool clinical populations, in this case, children diagnosed with Phenylketonuria. In their study, executive performance was related to Phenylalanine levels, which is considered an index of cortical dopaminergic activity in this population. Adequate measurement of these skills may be critically important, as executive abilities have been demonstrated to contribute to academic and behavioral difficulties in clinical populations, independent of general intellectual abilities (Taylor, Schatschneider, Petrill, Barry, & Owens, 1996). Particularly for young children, if executive abilities can be reliably assessed prior to school entry, early intervention could be provided to reduce the adverse impact of the particular disease or disorder on outcome.

Second, the structure and function of the prefrontal cortex changes significantly in the preschool period, including large-scale pruning of synaptic connections (Huttenlocher, 1979) and maturation of subcortical prefrontal myelination (Kinney, Brody, Kloman, & Gilles, 1988). Studies using resting EEG recordings (Thatcher, 1991, 1994) have identified a cycle of brain electrical signal development between 1 and 5 years of age, characterized by: (a) increased coherence in electrical activity between the short distance, anterior electrode recording sites, (b) lengthened fronto-lateral connections that became synchronous prior to frontal dorsomedial and central sites in the left hemisphere, and (c) lateral to medial differentiation of long-distance connections to shorter fibers in the right hemisphere. Because of these ongoing normative changes in the prefrontal cortex, the ability to measure developmental changes in executive skill during the preschool period is particularly important.

Which specific skills and tasks are defined as "executive" is a debated issue (Lyon & Krasnegor, 1996). In children, the contribution of working memory and inhibition recently has been emphasized (Roberts & Pennington,1996; Goldman-Rakic, 1987), although traditional, but more molar, skills, such as judgement and problem solving, also are included (Denkla, 1996). In the normative study of preschool children by Welsh, Pennington, and Grossier (1991), tasks specifically designed to measure working memory were not used. The animal neuroscience literature, where rich brain-behavior relations between the prefrontal cortex and discrete executive functions have been established, offers a source of tasks to be adapted for use in human children (Diamond & Goldman-Rakic, 1989; Kaufmann, Leckman, & Ort, 1989). The purpose of this study was to: (a) develop normative data for preschool children on tasks adapted from developmental and neuroscience literature, and (b) determine whether task performance differed by age group.

METHOD

Participants

Ninety-eight preschool children, aged 26 to 66 months ($M = 43.53$ months, $SD = 10.24$), were recruited from birth announcements and local preschools. Children were grouped in the following age group increments: 30 months ($n = 20$, $M = 30.71$, $SD = 2.95$, Range 26–34 months), 36 months ($n = 21$, $M = 36.98$, $SD = 1.51$, Range 35–39 months), 42 months ($n = 20$, $M = 42.97$, $SD = 1.06$, Range 41–44 months), 48 months ($n = 19$, $M = 49.48$, $SD = 2.49$, Range 45–52 months), and 60 months ($n = 18$, $M = 59.79$, $SD = 3.48$, Range 53–66 months). There were 53 males and 45 females, with comparable sex distribution across age groups (30-month $n = 10$ males, 36-month $n = 12$ males, 42-month $n = 12$ males, 48-month $n = 10$ males, 60-month $n = 9$ males; $\chi^2[4, N = 98] = 0.63, p > .95$). The majority of the sample ($n = 91$) were Caucasian. The average maternal education was 17.3 years ($SD = 2.1$). All children weighed more than 2500 grams at birth and developmental milestones reportedly were achieved appropriately.

Procedure

Children were tested individually in a quiet testing room. The examiner sat across from, but adjacent to, the child, at a small, low table. The entire executive function battery took about an hour (45 min for the younger children and 75 min for the older children).

Breaks were administered to maintain cooperation and interest. Testing was scheduled at times reported by parents not to interfere with regular naps or meals. All examiners were blind to the hypotheses of the study. The type of reward (i.e., small stickers, M & M® Baking bits, colored Rice Krispies®, raisins, Cheerios®, and pennies) was changed at the beginning of each delayed-response format task and again if the child appeared to be losing interest in order to maintain a high level of motivation.

Measures

A-not-B (AB; Diamond, 1988). AB was included because of its demonstrated relation to the dorsolateral prefrontal cortex in studies with animals, normally developing infants (Diamond, 1985), and children with clinical conditions (Diamond, Prevor, Callender, & Druin, 1997; Espy, Kaufmann, & Glisky, 1999). Researchers have suggested that two component skills are necessary to successfully complete the task: working memory and inhibition (Goldman-Rakic, 1987; Espy, Kaufmann, McDiarmid, & Glisky, 1999).

The child watched the examiner hide the reward in one of two shallow wells on a testing board. Both wells were covered simultaneously by two identical beige coffee cups. The testing board then was placed out of the child's sight under the table to prevent location cueing (e.g., leaning to the side of the reward) that has been demonstrated to improve performance (Diamond, 1985). The examiner counted aloud for 10 s in an engaging, melodic voice to maintain interest in the task and to distract the child from the testing board. A 10 s delay was chosen to maximize the number of children who completed the task, as it was observed during pilot testing that many children got up repeatedly from their seat and/or quickly lost interest in the tasks if longer delays were used. A constant delay was chosen, in order to maintain a consistent administration procedure across the delayed response format tasks. At the end of the 10-s delay, the testing board was returned to the table. The child then retrieved the reward by displacing the chosen cup. The child was allowed to keep or consume the reward only on correct trials. The reward was moved to the alternate well after the child retrieved the reward correctly for two consecutive trials on all subsequent trials. Ten trials were administered, as Espy, Kaufmann, McDiarmid and Glisky (1999) found that 10 AB trials was sufficient to elicit individual performance variability. An error was scored when the child firmly touched or began to lift the cup on the unrewarded well. If the child subsequently reached toward or displaced the correct cup, the first response still was considered an error, consistent with scoring procedures used by Diamond (1985). Five dependent variables were calculated: the number of correct retrievals (ABCORR), the number of correct consecutive responses (ABCRUN), the number of correct consecutive two trial sets achieved (ABSETS), the number of perseverative errors committed after the first correct two- trial set was achieved (ABPERR), and number of trials in the longest run of consecutive perseverative errors (ABPRUN).

Delayed Alternation (DA; Goldman, Rosvold, Vest, & Galkin, 1971). DA also was included because of the presumed reliance on working memory (Goldman et al., 1971) as DA performance has been linked to dorsolateral prefrontal cortical function in animals (Diamond, 1991; Goldman-Rakic, 1987) and loads with AB (Espy, Kaufmann, McDiarmid, & Glisky, 1999), although lesions to other brain areas (posterior parietal cortex) also affect performance (Diamond, 1990). In DA, the testing board with the two lateral wells and the two beige coffee cup covers also were used. Unlike AB, the reward was hidden out of the child's sight in DA. The child had to "discover" the hiding rule. To achieve the maximal correct, the child had to alternate retrieval between right and left wells on each successive trial after the 10-s delay. When the child disrupted the alternation by erroneously searching on the same side, the examiner hid the reward at the same location until correct retrieval occurred, thereby resuming the alternating sequence. Twenty trials were administered. Five dependent measures were scored, the number of correct responses (DACORR), the number of correct alternations (DACALT), the number of consecutive correct trials in the longest run of alternations (DACRUN), the number of perseverative errors (DAPERR), and number of trials in the longest perseverative run (DAPRUN). Because there were only two wells, all error responses were perseverative.

Spatial Reversal (SR; Kaufmann et al., 1989). SR was used primarily to measure shifting or cognitive flexibility, as the child had to flexibly shift among response sets. It was modeled after the object reversal tasks used in the animal neuroscience literature, where performance differences have been noted following lesions to the orbital frontal cortex (Mishkin, 1964). Reversal tasks do not share large variablity with AB or DA, suggesting less reliance on working memory or inhibition (Espy, Kaufmann, McDiarmid, & Glisky, 1999). SR also has been used with developmentally delayed populations (McEvoy, Rogers, & Pennington, 1993; Griffith, Pennington, Wehner, & Rogers, 1999). SR is similar to AB and DA, as all used the testing board and coffee cup covers. Like DA, the child did not observe the hiding of the reward. SR used a spatial rule for hiding. Unlike DA, but similar to AB however, the child retrieved the reward at the same location until a criterion of consecutive correct retrievals was met. In SR, the criteria was four consecutive correct trials, as pilot testing indicated that preschool children required more successful trials to establish the rule when they did not observe the hiding of the reward. After the child successfully retrieved the reward at a particular spatial location for four consecutive trials (i.e., a given lateral well), the reward then was hidden in the opposite lateral well. There were twenty trials, as the four-trial retrieval criterion required more trials in order to maintain a sufficient number of shifts between hiding locations. Four dependent measures were scored: number of correct responses (SRCORR), number of trials until the first correct set was achieved (SRFIRST), number of perseverative errors after the first correct set (SRPERR), and number of consecutive trials in the longest perseverative run (SRPRUN).

Color Reversal (CR; Kaufmann et al., 1989). Like SR, CR was postulated to measure shifting or cognitive flexibility, but SR and CR differed in the nature of the hiding rule. In SR, the rule was spatial; in CR, it was visual (color). Instead of using the beige coffee cups to cover the two wells, one blue and one yellow disc were used, where the colored discs moved between sides randomly across trials. As in SR and DA, the child did not observe the examiner hide the reward in CR. When the child retrieved the reward from beneath the colored disc correctly for four consecutive trials, the reward then was hidden beneath the disc of the other color. Twenty trials were administered. Four dependent measures were scored: number of correct responses (CRCORR), number of trials until the first set was achieved (CRFIRST), number of perseverative errors after the first correct set (CRPERR), and number of consecutive trials in the longest perseverative run (CRPRUN).

Self-Control (SC; Lee, Vaughn, & Kopp, 1983). This task was chosen to assess inhibition (Welsh & Pennington, 1988). In SC, the child was shown a reward. The examiner used an animated tone to comment on reward desirability (e.g., "These M&M's® sure look good. I like green ones, do you? Yum yum."). There were two trials. In SC1, the reward (M&M's®) was hidden under the beige coffee cup on the testing board (only a single well and cup were used). In SC2, the reward was a wrapped gift that was placed directly on the table. The child was instructed not to touch the reward while the examiner finished completing another task. The examiner then backed up from the table, turned partially away from the child, and reviewed test sheets while surreptitiously monitoring the child. The latency to touch the reward on each trial was scored with a maximum of 150 s.

Shape School (SS; Espy, 1997). This task also was included to measure inhibition. The Shape School is in a story-book format, depicting child-like figures in a school. It includes four conditions: Control, Inhibit, Switch, and Both, but only the Control and Inhibit condition are reported here because children younger than 48 months were not administered the Switch and Both conditions (see Espy, 1997). In the Control condition, the child was instructed to name the figure color in order, as fast as possible, without making any errors. In the Inhibit condition, the figures had two facial expressions, either happy or sad/frustrated. The child was instructed to name the figure color of the happy-faced figures and to inhibit naming or ignore the sad/frustrated-faced figures. There were 15 figures in each condition. The dependent measure for each condition was the number correct (SSCCORR, SSICORR), the time required to name all pertinent figures (SSCTIME, SSITIME). These scores then were transformed into efficiency scores (SSCEFF, SSIEFF; Efficiency = (the number of correct–the number of errors) / total time).

Tower of Hanoi (TOH; Welsh et al., 1991). This measure was included as a measure of molar problem solving and planning. Welsh and colleagues have used the TOH successfully in normally developing children (Welsh et al., 1991) and those with phenylketonuria (Welsh et al., 1990). The child moved three discs across pegs to achieve the model configuration on successively more difficult problems. With children in the age range of this study, an instructional story was used to describe the rules and goals of the task, involving three monkeys (rings) of different sizes (Daddy, Mommy, Baby) that may jump among trees (pegs). Unlike the administration in Welsh et al. (1991), each of the six problems (seven moves maximum) were presented for a maximum of two trials, in order to reduce task length. The dependent measure was the total number of problems solved in the minimal prescribed number of moves (TOHPS).

Intelligence The Peabody Picture Vocabulary Test-Revised, Form M (PPVT-R; (Dunn & Dunn, 1981) was administered to estimate intelligence in preschool children. The resulting standard score was the dependent measure. The sample mean PPVT-R standard score was 110.57 ($SD = 10.78$, Range 86–140), with no differences by age group, $F(4, 74) = 1.72$, $p > .15$).

Executive function task performance by age group is depicted in Table 1. There were children who were unable to complete particular tasks during administration ($n = 2$, DA; $n = 5$, SR; $n = 4$, CR; $n = 7$, SC2; $n = 6$ VS; $n = 5$ SS Control, $n = 7$ SS Inhibit). Three children inadvertently were not administered SC1 and SC2. Nineteen children did not complete the PPVT-R due to fatigue (the PPVT-R was administered at the end of the battery) or were not administered the task due to examiner error.

Design and Analysis

A multivariate analysis of variance design was used to examine age group-related performance. Separate MANOVA's were conducted for each task using the dependent measures within a task. If the Wilk's Lambda value associated with the age group effect in the overall MANOVA was significant, then the univariate ANOVA's for each task dependent variable were examined. Post hoc Tukey LSD pairwise comparisons were conducted to examine at what age group performance differed. Individual differences in task performance due to sex and intelligence also were examined. For these analyses, sex was used as a dichotomous independent variable in the pertinent MANOVA. Where sex significantly predicted task performance, it was re-entered in to another MANOVA with age group and the interaction of sex and age group, in order to examine whether the effect of sex moderated the age group effect. In order address the relation of intelligence and executive function task performance, correlations between PPVT-R standard score and each task dependent variable were examined after the effect of age was controlled statistically. All analyses were conducted with SAS Version 6.12.

RESULTS

Executive function task performance by age group is presented in Tables 1 and 2. Consistent with prediction, there were significant main effects of age group on the AB (Wilk's $\Lambda = .68, F[20, 296] = 1.86$, $p < .02$), DA (Wilk's $\Lambda = .61, F[16, 269] = 2.98, p < .001$), SS (Wilk's $\Lambda = .43$, $F[12, 167] = 5.28, p < .001$), and TOH ($F[4, 93] = 19.96$, $p < .001$) tasks. For AB, mean task performance differed by age group for all dependent measures, ABCORR ($F[4, 93] = 4.38, p < .01$), ABCRUN ($F[4, 93] = 5.79$, $p < .001$), ABSETS ($F[4, 93] = 4.89, p < .001$), ABPERR ($F[4, 93] = 6.53$, $p < .0001$), and ABPRUN ($F[4, 93] = 4.65, p < .01$). Examination of the mean performance on each AB dependent variable in Table 1 revealed steady, but relatively small, improvements in performance across age groups. In Table 3, Tukey comparisons among age groups revealed differences between younger (30-, 36-, and 42-month) and older (48- and 60-month) age groups for most AB dependent variables. Performance among 30-, 36-, and 42-month groups did not differ for all but one AB dependent variable (ABPERR). For all AB variables, performance was comparable between 48- and 60-month-old children.

A similar pattern of results was observed for DA. There were significant main effects of age group on DACORR ($F[4, 91] = 10.55$, $p < .001$), DACALT ($F[4, 91] = 10.19, p < .001$), DACRUN ($F[4, 91] = 8.71, p < .001$), DAERR ($F[4, 91] = 10.55, p < .001$), and DAPRUN ($F[4, 91] = 4.66, p < .002$). In Table 1, mean DA performance improved steadily, with age group differences larger than what was observed for AB. Tukey comparisons among age groups also were consistent with a more differentiated pattern of performance across age. A stepped pattern was observed, with performance on most DA variables differing significantly between the 30- and 36- and those older, and between the 42-month age group and those older. DA performance was comparable among 48- and 60-month-old children.

For SS, there were significant age group effects on the time variables, SSCTIME ($F[3, 66] = 10.42, p < .001$), and SSITIME ($F[3, 66] = 25.60$, $p < .001$). The number correct for the Inhibition condition differed marginally among age groups ($F[3, 66] = 2.49$, $p < .07$), but there were no age-related differences on SSCCORR. When the SS scores were transformed into efficiency scores, the age group effects in the overall MANOVA and univariate ANOVA's all were significant (overall Wilk's $\Lambda = .49, F[6, 130] = 9.22$, $p < .001$); SSCEFF $F[3, 66] = 8.43$, $p < .001$; SSIEFF $F[3, 66] = 20.93$, $p < .001$). Examination of the mean performance by age group in Table 2 revealed steady age-related

Table 1. Sample Task Performance by Age Group – Delayed Response Format Tasks.

	30 month			36 month			42 month			48 month			60 month		
Measure	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)
A-not-B															
CORR	20	7.75	(1.19)	21	7.86	(1.85)	20	8.00	(1.81)	19	9.00	(1.15)	18	9.22	(0.94)
CRUN	20	4.35	(2.03)	21	5.33	(2.74)	20	5.65	(2.51)	19	7.37	(2.69)	18	7.61	(2.61)
SETS	20	3.30	(0.73)	21	3.48	(1.12)	20	3.45	(1.15)	19	4.16	(0.90)	18	4.39	(0.78)
PERR	20	2.15	(1.14)	21	1.67	(1.31)	20	1.40	(1.10)	19	0.79	(1.03)	18	0.61	(0.70)
PRUN	20	1.20	(0.52)	21	1.38	(1.20)	20	1.20	(1.01)	19	0.58	(0.51)	18	0.50	(0.51)
Delayed Alternation															
CORR	19	10.00	(1.89)	20	11.30	(2.05)	20	12.59	(2.56)	19	13.67	(3.01)	18	14.57	(2.50)
CRUN	19	2.53	(1.26)	20	3.35	(1.09)	20	5.30	(3.38)	19	5.93	(3.14)	18	7.34	(4.20)
CALT	19	2.61	(1.95)	20	4.75	(2.47)	20	6.03	(3.83)	19	8.06	(4.65)	18	9.29	(4.34)
PERR	19	10.00	(1.89)	20	8.70	(2.05)	20	7.41	(2.56)	19	6.33	(3.02)	18	5.43	(2.50)
PRUN	19	2.60	(1.05)	20	2.35	(1.04)	20	1.91	(0.76)	19	1.80	(0.83)	18	1.54	(0.43)
Spatial Reversal															
CORR	20	13.63	(1.40)	18	13.50	(1.54)	20	13.15	(1.87)	18	13.77	(1.50)	17	14.05	(1.52)
FIRST	20	7.33	(4.47)	18	7.50	(4.57)	20	6.85	(4.04)	18	7.58	(3.99)	17	8.11	(4.18)
PERR[a]	19	5.02	(1.72)	18	4.72	(2.42)	19	5.58	(1.71)	17	4.52	(1.28)	17	4.06	(1.39)
PRUN	20	2.02	(0.70)	18	2.11	(1.13)	20	2.35	(0.93)	18	1.86	(0.86)	17	1.65	(0.86)
Color Reversal															
CORR	19	12.19	(2.10)	19	12.26	(2.25)	19	13.26	(1.52)	19	12.94	(1.74)	18	13.00	(1.64)
FIRST	19	11.98	(5.94)	19	12.63	(6.41)	19	9.79	(4.29)	19	10.73	(5.58)	18	9.44	(4.71)
PERR[a]	14	4.69	(2.04)	12	4.42	(1.24)	18	3.89	(1.78)	16	4.25	(1.81)	16	4.50	(1.59)
PRUN	19	4.46	(2.11)	19	3.74	(1.37)	19	3.42	(1.50)	19	3.26	(1.10)	18	3.50	(1.25)

Note. CORR = Number correct trials; CRUN = Maximal number of consecutive correct trials; SETS = Number of correct criterion sets; PERR = Number of Perseverative errors; PRUN = Maximal number of consecutive perseverative errors; CALT = Number of correct alternations; FIRST = Number of trials until criterion achieved. [a]The number of subjects is less for Color and Spatial Reversal PERR scores because perseverative errors are calculated after the first correct set is achieved. There was 4 subjects for SR and 18 subjects for CR who never achieved a correct four-trial set.

decreases in the latency to name all stimuli in both the Control and Inhibit conditions, and small increases in the number of stimuli correctly identified in the Inhibit condition. The number of stimuli correctly identified in the Control condition was close to ceiling, as 15 total stimuli were presented. Examination of post hoc Tukey age group comparisons in Table 3 revealed performance differences between 36-months-olds and older age groups on SS Control condition time and efficiency scores. For the SS Inhibit condition, 36-month-olds and those of older ages differed in mean SSI time and efficiency scores. Furthermore, mean SSITIME and SSIEFF scores differed between children 42 months of age and those older.

On TOH, the effects of age group also were significant. In Table 2, it is evident that mean TOHPS performance increased steadily with age. The mean number of problems solved differed between the younger (30-, 36- and 42-month) and older (48- and 60-month) age groups, with no differences among the younger or older age groups, respectively.

Contrary to prediction, there were no age group differences in SR, CR, or SC performance (all p's $> .49$). In a previous paper (Espy, Kaufmann, & Glisky, 1999), SC discriminated among cocaine-exposed and non-exposed toddlers when scored on a pass/fail (a pass score was obtained if the subject inhibited retrieving the M&M® reward or gift for the full 150 s). In Table 2, SC

Table 2. Sample Task Performance by Age Group – Other Executive Function Tasks.

	30 month			36 month			42 month			48 month			60 month		
Measure	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)
Self Control															
Condition 1															
TIME	20	125	(46.54)	21	132	(44.40)	20	137	(40.98)	16	146	(16.75)	18	150	(0.00)
Condition 2															
TIME	19	125	(46.95)	20	119	(51.18)	17	127	(45.13)	17	136	(38.99)	15	150	(0.00)
Tower of Hanoi															
Problems															
Solved	20	1.15	(1.18)	21	1.48	(0.87)	20	1.90	(1.33)	19	3.73	(0.99)	18	3.89	(1.81)
Shape School[a]															
Control Condition															
TIME	—			25	49.52	(29.78)	17	27.41	(7.43)	18	25.39	(8.98)	16	23.75	(9.88)
CORR	—			25	14.72	(0.61)	17	14.88	(0.48)	18	14.94	(0.24)	16	14.94	(0.25)
EFF	—			25	0.36	(0.14)	17	0.59	(0.21)	18	0.67	(0.29)	16	0.72	(0.26)
Inhibit Condition															
TIME	—			21	83.71	(38.79)	16	41.88	(20.14)	17	26.82	(10.86)	16	22.56	(7.02)
CORR	—			22	13.45	(2.22)	16	13.69	(2.21)	18	14.56	(0.70)	16	14.81	(0.54)
EFF	—			21	0.20	(0.16)	16	0.39	(0.26)	17	0.62	(0.23)	16	0.71	(0.21)

Note. TIME = time in s, CORR = Number correct, EFF = (# correct – # incorrect / time).
[a]The Shape School (Espy, 1997) was administered to children age 33 months and above.

Table 3. Age Group Differences on Each Task.[a]

Task	Significant differences among age groups					
A-not-B						
CORR	30–48	30–60	36–48	36–60	42–48	42–60
CRUN	30–48	30–60	36–48	36–60	42–48	42–60
SETS	30–48	30–60	36–48	36–60	42–48	42–60
PERR	30–42	30–48	30–60	36–48	36–60	42–60
PRUN	30–48	30–60	36–48	36–60	42–48	42–60
Delayed Alternation						
CORR	30–42	30–48	30–60	36–48	36–60	42–60
CRUN	30–42	30–48	30–60	36–42	36–48	36–60, 42–60
CALT	30–42	30–48	30–60	36–48	36–60	42–60
PERR	30–42	30–48	30–60	36–48	36–60	42–60
PRUN	30–42	30–48	30–60	36–48	36–60	42–60
Spatial Reversal						
PERR	42–60					
PRUN	42–60					
Self Control Condition 2						
TIME	36–60					
Tower of Hanoi						
PS	30–48	30–60	36–48	36–60	42–48	42–60
Shape School Control Condition						
TIME	36–42	36–48	36–60			
EFF	36–42	36–48	36–60			
Inhibit Condition						
TIME	36–42	36–48	36–60	42–60		
CORR	36–60					
EFF	36–42	36–48	36–60	42–48	42–60	

Note. CORR = Number correct; CRUN = Maximal number of consecutive correct trials; SETS = Number of correct criterion sets; PERR = Number of Perseverative errors; PRUN = Maximal number of consecutive perseverative errors; CALT = Number of correct alternations; FIRST = Number of trials until criterion achieved; TIME = time in s; PS = Problems Solved; EFF = (# correct – # incorrect/time).
[a]There were no significant age group differences for SRCORR, SRFIRST, CRCORR, CRPERR, CRPRUN, CRFIRST, SC1, SSCCORR.

variability across age groups differed considerably, consistent with performance ceiling in older children. Therefore, the SC data was reanalyzed, comparing the distribution of children who inhibited retrieving the reward for the entire observation period across age groups. In order to increase the number of subjects in each cell, groups were collapsed into 36, 48, and 60 months of age. However, because all children in the 60-month age group inhibited responding for the full 150 s, the 48- and 60-month age groups were collapsed further in order to avoid a cell with a zero count. There were significant differences in the number of children who inhibited, with the 36-month age group being less likely to inhibit responding than older children (Fischer's Exact = .04, χ^2 [1, N = 95] = 4.43). On SC1, there were 33 children in the 36-month age group who inhibited reaching for the M&M®, where there were 51 older children. A similar result, although marginally significant ($\chi^2[1, N = 88] = 3.55$, $p < .06$), was obtained for SC2, where only 28 children in the 36-month age group inhibited touching the gift, compared to 43 older children.

Task Performance Differences Related to Sex and IQ

Males and females performed comparably on DA, SR, CR, SC, TOH, and SS. However, sex was related marginally to AB performance (Wilk's

$\Lambda = .89$, $F[5, 92] = 2.20$, $p < .07$), but did not vary with respect to age group (Sex*Age Group Wilk's $\Lambda = .80$, $F[20, 279.55] = 0.97$, $p > .49$). Because AB performance differed marginally among males and females, the univariate ANOVA's examining the main effect of sex on performance were examined. There were significant univariate sex-related performance differences on all AB dependent variables. Female preschool children made more correct retrievals (Female $M = 8.71$, $SD = 1.65$; Male $M = 8.02$, $SD = 1.38$), retrieved the reward for more consecutive trials (Female $M = 6.89$, $SD = 2.91$, Male $M = 5.26$, $SD = 2.43$), obtained more correct two-trial sets (Female $M = 4.04$, $SD = 1.02$; Male $M = 3.47$, $SD = 0.97$), and made fewer perseverative errors (Female $M = 1.00$, $SD = 1.11$; Male $M = 1.64$, $SD = 1.21$) and consecutive perseverative runs (Female $M = 0.80$, $SD = 0.84$; Male $M = 1.15$, $SD = 0.89$) than males.

After statistically removing the effect of age, intelligence, as measured by the PPVT-R standard score, was not related to AB, DA, SR, CR, or SC performance. PPVT-R standard scores were related to performance on TOH ($r^2_{ppvtrss(tohps.age)} = .09, t[61] = 3.05, \quad p < .01$), SSITIME($\mathrm{r}^2_{\mathrm{ppvtrss(ssitime.age)}} = .04, \quad t[61] = -2.13$, $p < .04$) and SSCEFF ($r^2_{ppvtrss(ssceff.age)} = .05, t$ $[66] = 2.06, p < .05$).

DISCUSSION

In preschool children, performance on many of the executive function tasks improved across the 30- to 60-month age groups. On both AB and DA, older preschool children retrieved the reward on more trials and made fewer perseverative errors than did younger children. Therefore, AB and DA were sensitive to age group-related differences beyond the late infancy age range that has been more commonly studied (Diamond, 1985; Piaget, 1954; Wellman, Cross, & Bartsch, 1986). These findings are consistent with those of Diamond et al. (1997) who have found age-related change on AB in their normal control subjects. In the present study, greater age group-related differences were noted on DA, as many of the older children were performing at ceiling on AB. DA may be more difficult, and therefore, more appropriate across a wider age range. In studies using AB with younger populations, there are relatively large individual differences in the delay necessary to elicit perseverative responding (Bell & Fox, 1992; Diamond, 1985). In this study, perseverative errors may have been minimized in some children, as a 10-s delay may not have been sufficient to elicit AB errors for all children. In other studies with AB where a constant delay also was used (Diamond et al., 1997), reduced individual performance variability in older children also was observed. It is likely that using a longer delay with the older children would have elicited more AB errors. For the purpose of this study, however, it was considered necessary to maintain a constant delay to compare performance across age groups and tasks. Despite the constant delay in this study, performance still differed by age group.

Similar age group-related improvements were noted on the Shape School task. This task theoretically may provide a method to assess important executive function components, such as inhibition and shifting, relatively more independently. It also differs in format from the delayed response tasks, and yet allows investigation of similar constructs. Similar to Espy (1997), task performance was related to age group in the present study. In both the Espy and the present studies, 36-month-old children performed more poorly in the Control and Inhibit conditions relative to older children. Unlike findings from Espy, performance differed further between the 42-month and the older age groups. In the present study, age was broken down into 6-month age group segments allowing for a more detailed analysis. These findings demonstrate the importance of parsing age into small units during this period of rapid change.

Similar to Welsh et al. (1991), performance on the Tower of Hanoi differed among age groups, despite the shorter, two-trial administration format used in this study. This task appears to be a robust measure of preschool problem solving and has become an important tool to investigate executive functioning in adult and child clinical populations. A Tower of Hanoi variant is included on several newly developed neuropsychological measures, including the NEPSY (Korkman, Kirk,

& Kemp, 1998). In adults, Welsh, Satterlee-Cartmell, and Stine (1999) recently found that performance on Tower of Hanoi, Contingency Naming (Taylor, Albo, Phebus, Sachs, & Bierl, 1987) and Stroop (Golden, 1978) were correlated. This shared variability was interpreted to reflect shared reliance on working memory skills, and on inhibition, to a lesser degree. The cognitive underpinnings of Tower of Hanoi performance in preschool children is not known, but could shed light upon the neuropsychological changes in executive performance across the lifespan.

In contrast, SR and CR performance did not vary as consistently across age groups as the other delayed response format tasks, AB and DA. The reasons may be methodological, in that some of the youngest children were not able to complete the reversal tasks or never established a response set. For CR, these difficulties may be related to age-based skill differences in color identification. These measurement issues reduced task variability in the youngest children and may have reduced the power to detect age effects on these tasks. Reversal task performance also may be less reliable than AB in this age range. The study design did not permit investigation of this issue, but certainly, future studies should address reliability and validity issues. It is not clear why some of the youngest children performed adequately on AB but not on SR, as on the surface, both tasks appear to require similar skills to learn a spatial response set. The memory demand in SR may have been greater than in AB, due to the longer criterion and not observing the hiding of the reward. In fact, the number of criterion sets achieved for SR ($M = 2.01$, $SD = 0.88$) was less than that for AB ($M = 3.73$, $SD = 1.03$). Although both AB and SR require a spatial response, these tasks also do not load on the same factor (Espy, Kaufmann, McDiarmid, & Glisky, 1999), suggesting that they measure different cognitive skills.

Neuroanatomically, reversal task performance more often is linked to the ventral surface of the prefrontal cortex, in contrast to performance on AB and DA that typically is associated with the dorsolateral surface. This developmental pattern – minimal performance changes over the observed age range for some children, whereas other children were unable to achieve a response set or complete the task – suggests that the maturational course for reversal task performance is staged or stepped, rather than gradually unfolding across age. Because even young infants can complete AB, the development of the cognitive skills used in AB may represent increases in efficiency, rather than basic skill acquisition. In contrast, those skills required for reversal task performance might have a more protracted phase of acquisition during the preschool period, and therefore, the pattern of development would differ. A longitudinal design would better address this issue.

These measures, then, are useful for examining executive skills in preschool children. Performance on these measures was largely independent of intelligence, consistent with other studies examining executive skill development in older children (Levin et al., 1991; Welsh, Pennington, & Groisser, 1991). One of the prominent clinical signs of frontal lobe dysfunction in adults is gross deficits in judgement, planning, working memory, and inhibition, in light of preserved intellect (Eslinger & Damasio, 1985). The lack of relation between early verbal intelligence and executive test performance in normal preschool children is consistent with this skill discrepancy. These findings suggest that these executive function tasks measure something distinct from intelligence in preschool children. Theoretically, these findings also imply that even in preschool children, whose abilities are not as differentiated as those of adults or older children, intelligence does not capture the full range of neuropsychological skills. It will be important to determine whether these early executive skills also are related to later differences in outcome, such as academic achievement, as has been demonstrated in older children (e.g., Taylor, Schatschneider, Petrill, Barry, & Owens, 1996).

Furthermore, there were few sex-related differences in executive function task performance. On AB, females outperformed males on all dependent measures. These results are consistent with those of Diamond (Diamond, 1985; Diamond & Doar, 1989) where female infants tolerated longer delays and searched for the hidden reward at an earlier age than male infants. Interestingly, when performance in 23- to 25-month-old children also are studied, as in Espy, Kaufmann, McDiarmid, and Glisky (1999), sex differences emerged only for the longest consecutive run of correct AB responses. Although these present findings may represent bona fide sex-related differences particular to AB, they were not large in magnitude. Neither this study, nor those by Diamond, found that sex moderated the relation between age group and AB performance.

More generally, Diamond (1990) related delayed response performance to prefrontal cortical maturity in infancy. Clearly, there are important changes in the dorsolateral prefrontal cortex that occur later in the preschool years, mainly in synaptic reorganization (Huttenlocher, 1979; Thatcher, 1997). Whether the age-related changes on these executive function tasks used in this study reflect the continued maturation of the prefrontal cortex during the preschool period is an unanswered question. Overman, Bachevalier, Schuhmann, and McDonough-Ryan (1997) posited that performance discrepancies between reversal and concurrent discrimination tasks in young children reflect different maturational timetables of particular cortical areas. However, other researchers (e.g., Diamond, 1991) do not consider maturational differences in behavior to be isomorphic with the concurrent development of brain structure. Concurrent measures of brain function and executive function performance in preschool children is necessary in order to examine this issue directly. Such investigations remain difficult to conduct because brain measurement techniques with adequate spatial resolution used to assess these relations in adults, such as fMRI, are not yet well suited for use with normal preschool children. High-density array evoked potential recordings may offer a better alternative with this population.

Clinical neuropsychological investigations are just beginning to be conducted in preschool children, particularly those specifically examining executive functioning. For example, AB performance differed in prenatally cocaine-exposed toddlers compared to non-exposed controls (Espy, Kaufmann, & Glisky, 1999), in toddlers with Phenylketonuria relative to normals and controls (Diamond et al., 1997), and in infants born prematurely versus those born at term (Ross, Tesman, Auld, & Nass, 1992). AB, SR, and CR also have been used in older, but cognitively limited developmentally disabled populations, such as children with autism and mental retardation, with mixed results (Griffith et al., 1999; Kaufmann et al., 1989; McEvoy et al., 1993). Further investigations using these tasks with clinical populations will allow delineation of the nature of the disease process on cognition, but also will determine whether these tasks have discriminative validity. More generally, these findings suggest that tasks adapted from developmental and neuroscience paradigms offer a rich methodology by which to examine brain-behavior links in preschool children.

REFERENCES

Bell, M.A., & Fox, N.A. (1992). The relations between frontal brain electrical activity and cognitive development during infancy. *Child Development, 63*, 1142–1163.

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

Denckla, M.B. (1996). A theory and model of executive function: A neuropsychological persepctive. In G.R. Lyon & N.A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263–278). Baltimore: Paul A. Brookes Publishing Co.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development, 56*, 868–883.

Diamond, A. (1988). The abilities and neural mechanisms underlying AB performance. *Child Development, 59*, 523–527.

Diamond, A. (1990). The development and neural bases of memory functions as indexed by AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences, 608*, 267–317.

Diamond, A. (1991). Guidelines for the study of brain-behavior relationships during development. In H.S. Levin, H.M. Eisenberg, & A.L. Benton, (Eds.), *Frontal lobe function and dysfunction* (pp. 339–380). New York: Oxford University Press.

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology, 22*, 271–294.

Diamond, A., & Goldman-Rakic, P.S. (1989). Comparison of human infants and rhesus monkeys on Paiget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research, 74*, 24–40.

Diamond, A., Prevor, M.B., Callender, G., & Druin, D.P. (1997). Prefrontal cortex cognitive deficits in children treated continuously for PKU. *Monographs of the Society for Research in Child Development, 62*, 1–205.

Dunn, L.M., & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test–Revised (Form M)*. Circle Pines, MN: American Guidance Service.

Eslinger, P.J., & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35*, 1731–1741.

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

Espy, K.A., Kaufmann, P.M., & Glisky, M.L. (1999). Neuropsychologic function in toddlers exposed to cocaine in utero: A preliminary study. *Developmental Neuropsychology, 15*, 447–460.

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.

Golden, C.J. (1978). *Stroop Color Word Test*. Chicago: Stoetling Press.

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.

Goldman-Rakic, P.S. (1987). Development of cortical circuitry and cognitive function. *Child Development, 58*, 601–622.

Griffith, E.M., Pennington, B.F., Wehner, E.A., & Rogers, S.J. (1999). Executive functions in young children with autism. *Child Development, 70*, 817–832.

Heaton, R.K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources.

Huttenlocher, P.R. (1979). Synaptic density in human frontal cortex–developmental change and effects of aging. *Brain Research, 163*, 195–205.

Kaufmann, P., Leckman, J.M., & Ort, S.I. (1989). Delayed response performance in males with Fragile-X. *Journal of Clinical and Experimental Neuropsychology, 12*, 69.

Kinney, H.C., Brody, B.A., Kloman, A.S., & Gilles, F.H. (1988). Sequence of central nervous system myelination in human infancy. *Journal of Neuropathology and Experimental Neurology, 47*, 217–234.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A Developmental Neuropsychological Assessment*. San Antonio, TX: The Psychological Corporation.

Lee, M., Vaughn, B.E., & Kopp, C.B. (1983). The role of self-control in young children's performance on a delayed response memory for location task. *Developmental Psychology, 19*, 40–44.

Levin, H.S., Mendelsohn, D.B., Lilly, M.A., Fletcher, J.M., Culhane, K.A., Chapman, S.B., Harward, H., Kusnerik, L., Bruce, D., & Eisenberg, H.M. (1994). Tower of London performance in relation to magnetic reasonance imaging following closed head injuries in children. *Neuropsychology, 8*, 171–179.

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

Lyon, G.R., & Krasnegor, N.A. (1996). *Attention, memory, and executive function.* Baltimore: Paul H. Brookes Publishing Co.

Mattson, S.N., Goodman, A.M., Caine, C., Delis, D.C., & Riley, E.P. (1999). Executive functioning in children with heavy prenatal alcohol exposure. *Alcoholism: Clinical and Experimental Research, 23*, 1808–1815.

McEvoy, R.E., Rogers, S.J., & Pennington, B.F. (1993). Executive function and social communicative deficits in young autistic children. *Journal of Child Psychology and Psychiatry, 34*, 563–578.

Mishkin, M. (1964). Perseveration of central sets after frontal lesions in monkeys. In J.M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 219–241). New York: McGraw-Hill.

Overman, W.H., Bachevalier, J., Schuhmann, E., McDonough-Ryan, P. (1997). Sexually dimorphic brain-behavior development: A comparative perspective. In N.A. Krasnegor, G.R. Lyon, & P.S. Goldman-Rakic (Eds.), *Development of the prefrontal cortex: Evolution, neurobiology, and behavior* (pp. 85–113). Baltimore, MD: Paul H. Brookes.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

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

Ross, G., Tessman, J., Auld, P., & Nass, R. (1992). Effects of subependymal and mild intraventricular lesions on visual attention and memory in premature infants. *Developmental Psychology, 28*, 1067–1074.

Smith, A. (1983). Overview or "underview?" Comment on Satz and Fletcher's "Emergent trends in neuropsychology: An overview." *Journal of Consulting and Clinical Psychology, 51*, 768–775.

Taylor, H.G., Albo, V.C., Phebus, C.K., Sachs, B.R., & Bierl, P.G. (1987). Postirradiation treament outcomes for children with Acute Lymphocytic Leukemia: Clarification of risks. *Journal of Pediatric Psychology, 12*, 395–411.

Taylor, H.G., Schatschneider, C., Petrill, S., Barry, C.T., & Owens, C. (1996). Executive dysfunction in children with early brain disease: Outcomes post Haemophilus influenzae meningitis. *Developmental Neuropsychology, 12*, 35–51.

Thatcher, R.W. (1991). Maturation of the human frontal lobes: Physiological evidence for staging. *Developmental Neuropsychology, 7*, 397–419.

Thatcher, R.W. (1994). Psychopathology of early frontal lobe damage: Dependence on cycles of development. *Development and Psychopathology, 6*, 565–596.

Thatcher, R.W. (1997). Human frontal lobe development: A theory of cyclical cortical reorganization. In N.A. Krasnegor, G.R. Lyon, & P.S. Goldman-Rakic (Eds.), *Development of the prefrontal cortex: Evolution, neurobiology, and behavior* (pp. 85–113). Baltimore, MD: Paul H. Brookes.

Wellman, H.M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. In R.N. Emde (Ed.), *Monographs of the Society for Research in Child Development* (# 214, Vol. 51, pp. 67). Chicago, IL: University of Chicago Press.

Welsh, M.C., & Pennington, B.F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Psychology, 24*, 199–230.

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.

Welsh, M.C., Pennington, B.F., Ozonoff, S., Rouse, B., & McCabe, E.R.B. (1990). Neuropsychology of early-treated phenylketonuria: Specific executive function deficits. *Child Development, 61*, 1697–1713.

Welsh, M.C., Satterlee-Cartmell, T., & Stone, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition, 41*, 231–242.