11-1-2004

Using path analysis to understand executive function organization in preschool children

T. E. Senn

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

P. M. Kaufmann
University of Nebraska-Lincoln

Using Path Analysis to Understand Executive Function Organization in Preschool Children

Theresa E. Senn, Kimberly Andrews Espy,
and Paul M. Kaufmann

Department of Family and Community Medicine
Southern Illinois University School of Medicine

There continues to be no consensus definition of executive functions. One way to understand different executive function components is to study abilities at their emergence, that is, early in development, and use advanced statistical methods to understand the interrelations among executive processes. However, to fully determine the constructs of interest, these methods often require complete data on a large battery of tasks, which are difficult to obtain with young children. Path analysis is an alternative statistical technique that requires only a single measure of each construct, yet still allows researchers to investigate complex relations among measures, to compare nested models, and to compare model fit across groups. Therefore, 117 preschool children (ages 2 years 8 months to 6 years 0 months) completed several executive function tasks. Path analysis was used to determine the relations between complex problem solving and working memory, inhibition, and set shifting processes. The best-fitting model included paths from working memory and inhibition to problem solving, and a correlation between working memory and inhibition. Interestingly, in younger children, inhibition was the strongest predictor of problem solving, whereas working memory contributed more strongly in older children. Suggestions for useful statistical methods to investigate the relations among executive functions in children are discussed.

The nature of executive functions and the relations among executive processes are far from resolved (Lyon & Krasnegor, 1996), with Borkowski and Burke (1996) concluding that "the greatest problem hindering research on executive functioning

(EF) is a failure to find consensus on a general definition of the construct. ... "
Cognitive neuropsychological approaches often have focused on more microlevel
components, such as working memory and response inhibition (e.g.,
Goldman-Rakic, 1987; Roberts & Pennington, 1996). In contrast, approaches
based on data from clinical patients often include more macrolevel constructs,
such as social judgment, self-regulation, planning, and problem solving (e.g.,
Damasio & Anderson, 1993).

One promising development has been the attempt to integrate these micro-
and macrolevel approaches, by examining the relations among these putative ex-
ecutive functions. These micro and macro approaches were developed from two
different historical traditions and may reflect differing perspectives by which to
view similar phenomena. For example, cognitive neuropsychological researchers
view perseveration as an inherent by-product of breakdown in working memory
and inhibition processes (e.g., Roberts & Pennington, 1996). In clinical neuro-
psychological approaches, lack of flexibility can be seen as a fundamental deter-
minant of executive dysfunction in the social arena (e.g., Damasio & Anderson,
1993). One method of using both micro and macro approaches entails examining
the relations among microlevel variables (working memory, inhibition, and shift-
ing) and investigating the impact that these functions have on more macro abili-
ties, such as judgment, planning, and problem solving.

Previous studies have investigated relatively simple relations among these mi-
cro- and macrolevel executive functions. For example, in a sample of college stu-
dents, Miyake et al. (2000) used structural equation modeling to examine the con-
tributions of working memory, flexibility, and inhibition as predictors of
performance on a prototypic EF task, the Tower of Hanoi (TOH; Simon, 1975). In
TOH, the participant has to move rings on pegs in an initial state to reproduce the
configuration of a model, the end-goal state. The initial state of the rings deter-
mines the degree of difficulty of the problem, whereas the reproduction of the
model in the end-goal state can be achieved from as little as two moves for a rela-
tively simple problem to as many as 7, 11, or more moves for a more complex
problem. On the basis of face validity, TOH often is considered a planning or prob-
lem-solving task in the clinical neuropsychological literature (Lezak, 1995),
where subjects plan and implement move sequences of progressive complexity to
achieve the end-goal state. TOH performance is impaired in patients with damage
to the frontal lobe (Goel & Grafman, 1995; Levin et al., 1994) and in those with al-
tered dopaminergic neurotransmission (Owen, Doyon, Petrides, & Evans, 1996;
Welsh, Pennington, Ozonoff, Rouse, & McCabe, 1990). Several lines of evidence,
however, suggest that TOH can be better conceptualized as an inhibitory task. By
analyzing the strategy used to solve the TOH puzzles, Goel and Grafman found
that patients with diverse damage to the prefrontal cortex had difficulty making
moves that were counterintuitive, those that required backward movement in con-
flict with the longer term goal of peg configuration. Goel and Grafman interpreted

this difficulty as consistent with a lack of inhibition of prepotent (and more tempo-
rally salient) moves. In contrast, these patients did not have problems maintaining
move "stacks," or sequences of moves, therefore suggesting a minimal contribu-
tion of working memory. Using a different approach, Welsh, Satterlee-Cartmell,
and Stine (1999) compared the pattern of correlations between diverse executive
function tasks and TOH performance in normal adults. TOH performance was re-
lated to errors on the Contingency Naming Task, Part B (Taylor, Albo, Phebus, &
Sachs, 1987), and the Stroop Color and Conflict Scores (Golden, 1978), both of
which are inhibitory type tasks, but it was unrelated to measures of working mem-
ory, such as Spatial Working Memory Test (Owen, Downes, Sahakian, Polkey, &
Robbins, 1990) or Visual Memory Span (Wechsler, 1987). Therefore, Miyake et
al. (2000) compared a model with paths from the three hypothesized latent execu-
tive functions (flexibility, working memory, and inhibition) to TOH performance,
to a model with only a path from inhibition to TOH. In both models, correlations
among working memory, flexibility, and inhibition were included. The fit of the
model with one path from inhibition to TOH was equivalent to that of the full
model; therefore, the simpler, inhibition-only model was preferred.

These results highlight the utility of a latent variable approach and the impor-
tance of considering multiple levels of executive functions to determine the true
nature of executive function organization. Such layered approaches ultimately
may provide a better understanding of a wider array of cognitive functions, their
interrelations, and how they are disturbed by diverse neuropathology. These ap-
proaches have been underused in the neuropsychology (Francis, Fletcher, &
Rourke, 1988) and cognitive neuroscience fields generally, perhaps related to the
statistical sophistication, the diverse executive function battery necessary to fully
identify latent constructs, and the large number of participants required. In particu-
lar, obtaining complete data on large test batteries in clinical populations can be
difficult, due to a shortened attention span, fatigue, and an inability to understand
directions on some tasks. Such problems also are evident in children with neuro-
logical, psychiatric, and developmental disorders and normally developing young
children. In these populations, the simpler approach of path analysis could be used
to better understand cognitive organization. Path analysis can be considered a spe-
cial case of structural equation modeling, where observed performance on a speci-
fied test fully measures the latent construct of interest. Although this method does
not take advantage of the shared variance among multiple measures to define the
cognitive construct of interest, potential complex relations among multiple con-
structs can be tested using only one measure per construct. Importantly, the incre-
mental fit of nested path-analysis models can be compared directly by removing
one or more paths in the same manner that nested structural equation models are
compared. If the chi-square values of the two models do not differ, the simpler
model with fewer paths is preferred due to the principle of parsimony. Finally, path
analysis can be used to compare cognitive organization across multiple groups to

determine whether the same model fits the observed data from different groups equally well.

The purpose of this study was to illustrate how path analysis can be used to test hypothesized relations among executive functions in a population in which it is difficult to obtain complete data on multiple measures, that is, preschool children. A secondary purpose was to explore the nature of the organization among executive function constructs in preschool children. A multiple-group comparison analysis, using path analysis to compare model fit between younger and older children, also is illustrated. Models included both micro (i.e., working memory, inhibition, and flexibility) and macro (i.e., problem solving) executive functions to substantively address the relation among executive processes in this age range.

METHOD

Participants

Preschool children (ages 2 years 8 months to 6 years 0 months; M_{age} = 4 years 2 months, SD_{age} = 10 months) were recruited from birth announcements, local preschools, and the local health department. Among those preschool children participating in an ongoing cross-sectional study of executive function development (e.g., Espy, Kaufmann, Glisky, & McDiarmid, 2001; Espy, Kaufmann, McDiarmid, & Glisky, 1999), there were 117 children (60 boys, 57 girls) with complete data on all of the measures used here. The majority of the sample was White (n = 97), and the average maternal education of the sample was 15 years 8 months (SD = 2 years 8 months).

Procedure

The children were tested individually in a quiet room. The examiner, a trained graduate student, sat across from the child at a low table. A parent (usually the mother) was present in the room, completing questionnaires while the full battery of executive function tasks was administered, of which four tasks were selected for analysis here. The entire battery took approximately 60 to 90 min. A fixed task order was used, where tasks of different formats and cognitive demands were intermingled, to engage the child's maximal interest over the entire battery. The large number of tasks in the full battery and the need to separate tasks of similar format (e.g., tasks that used cups to hide rewards) prohibited counterbalancing. For a full description of the procedures and some of the other measures administered, see Espy et al. (2001, 1999).

Measures

Delayed alternation (DA; Espy et al., 1999).   DA was adapted from non-human animal neuroscience investigations (e.g., Goldman, Rosvold, Vest, & Galkin, 1971) for use with preschool children, as it has well-demonstrated relations to the dorsolateral prefrontal cortex and is considered a measure of working memory because a child holds information from a previous trial online to guide correct responding on the subsequent trial. In DA, a reward (small stickers, M&M Baking Bits®, colored Rice Krispies®, raisins, Cheerios®, or pennies) was hidden in one of two shallow wells on a testing board by the examiner, out of the child's sight. Both wells were covered by identical beige coffee cups. A pretrial was administered where neither well was baited, to determine the side of hiding on the test trials. After a 10-sec delay, the child retrieved the reward by displacing one of the cups, keeping the reward only if the correct cup were displaced. After correctly retrieving the reward, the reward was hidden in the opposite well. To obtain maximal rewards, the child had to search at alternating wells on each successive trial for 16 trials. The number of correct retrievals was scored.

Shape School (SS; Espy, 1997).   The SS is a Stroop-like task in a story-book format that includes different conditions designed to assess different executive functions. Here, the Inhibit condition was used to assess inhibition. First, the prepotent response was set up by having children name the color of 15 neutral-faced figures that are in line for a "school" activity. Then in the SS Inhibit condition, the child was presented again with the colored figures, some of whom had completed their work and were ready for the next activity (depicted with happy faces) and some who had not, shown with sad or frustrated faces. The child was instructed to name the color of the happy-faced figures, but not to name the colors of the sad-faced figures (to inhibit naming the figure color). An inhibition efficiency score was calculated (Efficiency = number correct/total time).

Spatial reversal (SR; Kaufmann, Leckman, & Ort, 1989).   Reversal tasks are prototypical flexibility tasks in the nonhuman animal neuroscience literature (Mishkin, 1964), which are believed to rely on ventromedial–orbitofrontal circuits for adequate performance (Watanabe, Kodama, & Hikosaka, 1997). In SR, the reward was hidden out of the child's sight, and the same wells used in DA were covered with beige coffee cups. First, there was a pretrial, where both wells were baited with a reward to establish the side of hiding for subsequent trials. Then, during the test trials, the reward was hidden in the same well (the side chosen during the pretrial) until the child correctly retrieved the reward on four consecutive trials; then the reward contingency was switched to the other well until the next consecutive four-trial set was achieved, where the contingency was reversed again. Sixteen

trials were administered. An efficiency score was calculated, the number of correct four-trial sets divided by the total number of errors. SR was administered before DA for all children because the child initially experienced success on the pretrial, regardless of choice, and because it was hypothesized to be easier to search at a previously rewarded location consistently. There were four other tasks that were administered between SR and DA to minimize any carryover effects.

Tower of Hanoi (TOH; Welsh, Pennington, & Groisser, 1991).   TOH was chosen as a macro-outcome variable because it is a more complex task than DA, SS, or SR and is considered to require multiple abilities, including multiple-executive functions (Bull, Espy, & Senn, 2004; Pennington, Bennetto, McAleer, & Roberts, 1996). TOH typically is considered to measure problem solving or planning ability (e.g., Simon, 1975). Because preschool children do not evidence overt behaviors that typically are considered to reflect planning (e.g., pausing before implementing move sequences, implementing organized move sequences), TOH likely taps problem solving to a greater degree than planning in this age range. TOH is composed of three discs that fit over three pegs; the child must move the discs across the pegs according to certain rules to achieve a model configuration. A story was used to describe the rules and goals of the task, involving three monkeys (rings) of different sizes (Daddy, Mommy, and Baby) that jump among trees (pegs) (Klahr & Robinson, 1981). Six successively more difficult problems were administered, where the number of moves to solve the problem increased progressively. A child was given 2 points for correctly solving the two-move problem, 3 points for correctly solving the three-move problem, and so on. If the child solved the problem using the fewest possible number of moves, a bonus of 25% of the points was given for that problem (e.g., if the child correctly solved the two-move problem in two moves, the child's score was $2 + .25 \times 2 = 2.5$ points). A total TOH score was obtained by summing scores across the six TOH problems.

Design

The model included four constructs, working memory, inhibition, and mental flexibility, as predictors of problem solving. Each construct was defined by a single task: working memory by DA, inhibition by SS, mental flexibility by SR, and problem solving by TOH, based on extant animal neuroscience literature or previous studies or both. For example, results of animal neuroscience investigations suggest that performance on DA and reversal tasks depends on different prefrontal circuits (Goldman et al., 1971; Mishkin, 1964), and in preschool children these tasks load on different factors (Espy et al., 1999), with differing trajectories of age-related differences in task performance (Espy et al., 2001). Although preliminary in nature, performance on the SS Inhibit condition loads with the NEPSY—A Developmental Neuropsychological Assessment (Korkman, Kemp, & Kirk, 2001) Statue subtest (Cwik, 2002), and it has a different pattern of development in pre-

school children from a task hypothesized to rely on flexibility (Espy, 1997). Although a single index was chosen to assess each construct, as with most executive function tasks, performance likely is not determined by a single, discrete ability. For example, in addition to requiring inhibition, SS also likely demands a degree of working memory proficiency to hold the instructions in mind while successfully inhibiting the prepotent stimuli. Here, however, these measures (DA, SS, SR) were selected as primarily representing the microconstruct of interest, where these three microcomponents have been found in both exploratory (e.g., Espy et al., 1999; Hughes, 1998) and confirmatory (e.g., Miyake et al., 2000) factor analysis studies of executive functions.

Given that TOH is presumed to involve higher order planning or problem solving, and therefore depends on multiple abilities, including working memory, inhibition, and flexibility (Bull et al., 2004), a complex model of associations among these constructs may better predict performance than a model with single, direct effects from working memory, inhibition, and flexibility to problem solving (e.g., Miyake et al., 2000). In the complex model, flexibility was hypothesized to act as a mediator between working memory, inhibition, and problem solving. Diamond (1988) hinted at the potential mediating role of flexibility when she theorized that dual disturbances in working memory and inhibition may lead to perseveration (the inverse of flexibility). In other words, if an individual is not able to maintain information over time or inhibit prepotent responses or both, he or she will continue to inflexibly choose the incorrect response. This lack of flexibility in turn is related to problem solving by limiting the ability to flexibly choose among alternate problem-solving strategies. Direct effects also were hypothesized from working memory and inhibition to problem solving (full model; see Figure 1). Path coefficients, which indicate the strength and direction of the relations between the constructs, are included in the figure. The full model is similar to the model tested by Miyake et al. (2000), as a breakdown in either working memory or inhibition was postulated to impair problem solving. Finally, given the competitive balance between working memory and inhibition proposed by Roberts and Pennington (1996), a correlation between working memory and inhibition was posited. This model was just identified; that is, based on the number of measures there were 10 possible variances and covariances that could be estimated, and in this model 10 variances and covariances were estimated (five paths between constructs, one correlation, and four residual variances), resulting in 0 df for the model and a perfect model fit (see Loehlin, 1998, for a more thorough discussion of model identification).

However, simpler models may account for the observed data as well as more complex models. Therefore, several nested models were tested, where one or more paths were removed, while retaining the same structure in the rest of the model. The incremental fit of the nested models was determined by calculating the difference in chi-square values and degrees of freedom between the two models. If the difference was not significant at the critical p value, then the more parsimonious
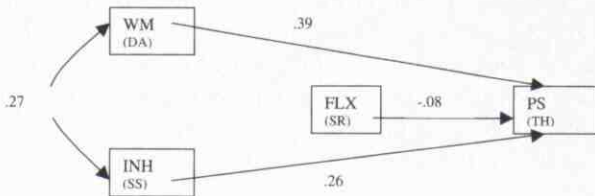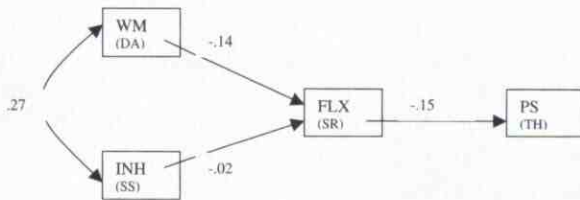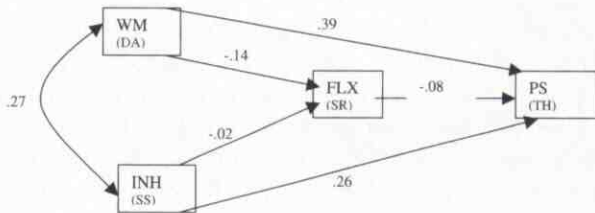
FIGURE 1   Top: Model 1, Full Model. Middle: Model 2, Flexibility as Mediator Only Model. Bottom: Model 3, Direct Paths Model.

model would be preferred. If the incremental difference in fit was significant, then the model with fewer paths fit the observed data more poorly and would be rejected. A model would not be considered nested if new, additional paths were added, and therefore differences in fit could not be compared by the difference in chi-square values and degrees of freedom. A theory-based and empirically driven rationale was used to develop all models.

In this vein, Model 2 tested whether the effects of working memory and inhibition on problem solving operated only through flexibility. Therefore, flexibility

still was included as a mediator in Model 2, but the direct paths from working memory and inhibition to problem solving were dropped (see Figure 1), making it a nested model compared to the full model. Model 3 was similar to that of Miyake et al. (2000), including only direct paths from working memory, inhibition, and flexibility to problem solving, and a correlation between working memory and inhibition (see Figure 1). This model was nested in comparison to the full model, as the paths from working memory and inhibition to flexibility were dropped. However, this model was not nested within Model 2 because this model included additional paths from working memory and inhibition to planning. Therefore, the incremental fit could not be tested directly against Model 2. Unlike Miyake et al. (2000), working memory and inhibition were not hypothesized to correlate with flexibility, as the premise of the full model is that flexibility mediates the contributions of working memory and inhibition on problem solving.

Some researchers have postulated that only working memory and inhibition are central executive functions (e.g., Goldman-Rakic, 1987; Roberts & Pennington, 1996). In Model 4, the paths to and from flexibility were removed, leaving only the direct paths from working memory and inhibition to problem solving. The correlation between working memory and inhibition also was retained (see Figure 2).
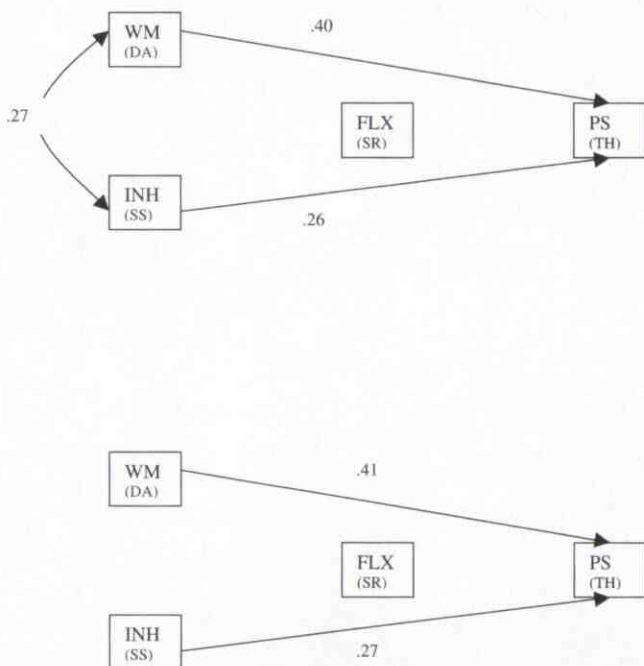

FIGURE 2   Top: Model 4, Working Memory & Inhibition Paths + Correlation Model. Bottom: Model 5, Working Memory & Inhibition Paths Only (independent) Model.

Although there is both theoretical (e.g., Roberts & Pennington, 1996) and empirical (e.g., Diamond, 1988; Espy et al., 1999) support for a correlation between working memory and inhibition, these executive functions may not be correlated, consistent with results by Hughes (1998) from a three-factor exploratory factor analysis. Therefore, to test the importance of this correlation, Model 5 included only direct paths from working memory and inhibition to problem solving without a correlation between working memory and inhibition (see Figure 2).

Because there is recent evidence (Miyake et al., 2000; Welsh et al., 1999) that inhibition is the primary contributor to TOH performance, at least in adults, Model 6 included only a direct path from inhibition to problem solving and a correlation between working memory and inhibition (see Figure 3). In contrast, other researchers have emphasized the centrality of working memory in solving the TOH (e.g., Simon, 1975) and in executive functions more generally (Kimberg & Farah, 1993; Pennington, 1997). In solving the TOH, children may work out the steps necessary to achieve the final goal mentally, placing demands on working memory. Therefore, Model 7 included only a direct path from working memory to problem solving, and a correlation between working memory and inhibition (see Figure 3).
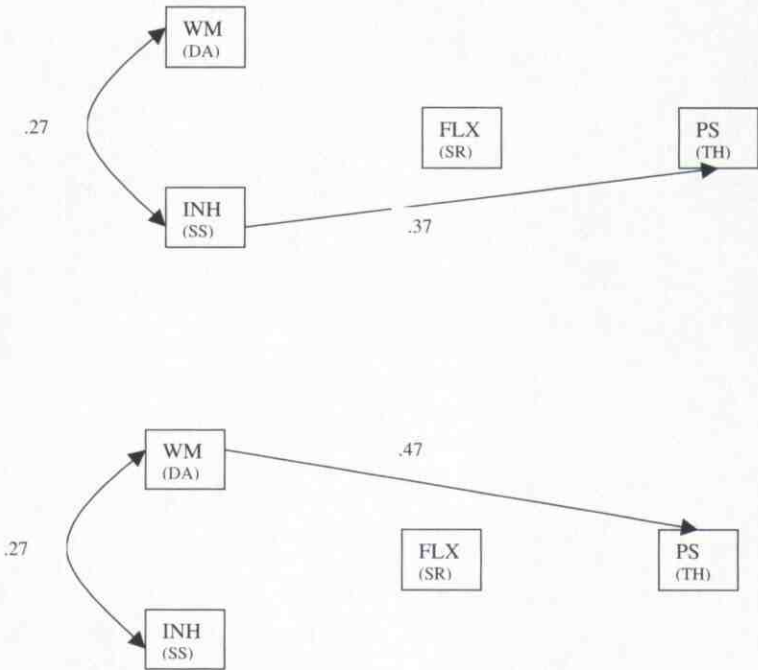

FIGURE 3   Top: Model 6, Inhibition Only Model. Bottom: Model 7, Working Memory Only Model

There are age-related differences in performance noted on some of the executive function tasks (Espy et al., 2001, 1999). Such performance differences may arise because of changes in proficiency. However, performance differences may not be a consequence of changes in proficiency, but rather differences in the nature of organization of executive control in children of different ages. Therefore, specifically testing different models of executive organization across children of varying ages can be informative. Here, an exploratory, multiple-group comparison was conducted to test whether executive function organization differed as a function of age group. The fit of the final model was tested against the observed data from children $\leq$ age 4 ($n = 46$) and that from children $>$ age 4 ($n = 71$). First, a multiple-group model was run in which path coefficients were allowed to differ across the two groups. Then, the same structural model was run in which path coefficients were constrained to be equal across the two age groups. If the latter model fit the observed data more poorly, as determined by the incremental difference in chi-square values and degrees of freedom, then the model would be rejected, as it would not fit the observed data equally well for younger and older children.

These path analyses were conducted using Mplus version 2.1 (Muthen & Muthen, 1998), using maximum likelihood estimation procedures. The advantage of using Mplus or other structural equation modeling programs over conventional regression analyses (e.g., Pedhazur & Schmelkin, 1991), is that fit indexes are computed for the model as a whole, rather than providing path coefficients only. All variables had sufficient variability and appeared to be distributed approximately normally.

To determine model fit, several different goodness-of-fit indexes were used, as recommended by Hu and Bentler (1999). The chi-square, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI) were used, where models with acceptable fit were defined as those with a nonsignificant chi-square (although caution must be used when interpreting chi-square, as it is highly dependent on sample size), with an RMSEA $< .06$, and a CFI $\geq .95$ (Hu & Bentler, 1999). In addition, the models yield a percentage of variance that is accounted for in the dependent construct or variable (TOH) by the predictor constructs or variables. The residual variance—that not accounted for in the model—contains measurement error, subject-related error, and the influences of variables not included in the model.

RESULTS

Mean scores and standard deviations on each of the dependent measures are presented in Table 1. Note the psychometric characteristics of the measured task performance, where no performance floors or ceilings were observed, consistent with our previous reports (Espy, 1997; Espy et al., 2001). Goodness-of-fit indexes are

TABLE 1
Group Performance on Measured Variables

Variable	All Participants[a]		Age ≤ 4[b]		Age > 4[c]	
	M	SD	M	SD	M	SD
DA	9.94	2.31	8.97	1.93	10.57	2.33
SS	0.54	0.31	0.35	0.23	0.61	0.29
SR	0.39	0.27	0.42	0.30	0.37	0.26
TOH	13.66	8.78	8.61	5.24	16.93	9.09

Note.   DA = Delayed Alternation number of correct retrievals; SS = Shape School-Inhibition Condition efficiency score; SR = Spatial Reversal efficiency score; TOH = Tower of Hanoi total score.
[a]$N = 117$. [b]$n = 46$. [c]$n = 71$.

presented in Table 2. As can be seen in this table, the data were an acceptable fit for several of the models (i.e., Model 1, Model 3, and Model 4; Hu & Bentler, 1999). However, the real value of the approach demonstrated here lies in the direct testing of nested models to determine which competing models provide the best fit to the observed data. The incremental fit provides a more direct test of elements of model fit, with less of a reliance on the more qualitative goodness-of-fit indexes. A comparison of the chi-square values for the models is presented in Table 3.

TABLE 2
Goodness-of-Fit Indexes

Models	χ^2	RMSEA	CFI
1	$\chi^2(0) = 0.00$	.00	1.00
2	$\chi^2(2) = 37.46$	.39	0.06
3	$\chi^2(2) = 2.65$	.05	0.98
4	$\chi^2(3) = 3.62$	.04	0.98
5	$\chi^2(4) = 12.49$	.14	0.76
6	$\chi^2(4) = 25.93$	.22	0.52
7	$\chi^2(4) = 13.62$	.14	0.80

Note.   RMSEA = root mean square error of approximation; CFI = comparative fit index.

TABLE 3
Comparison of Nested Models

Model Comparison	χ^2 Difference	df Difference	p Value
Model 2 versus Model 1	37.46	2	< .001
Model 3 versus Model 1	2.65	2	.27
Model 4 versus Model 3	0.97	1	.32
Model 5 versus Model 4	8.87	1	.003
Model 6 versus Model 4	22.31	1	< .001
Model 7 versus Model 4	9.64	1	.002

The difference in fit between Model 2 (flexibility as mediator only) and the full model was significant, indicating that the flexibility as mediator model did not adequately reproduce the observed data as well as the full model. The fit of Model 3 (only direct paths from working memory, inhibition, and flexibility) also was compared to that of the full model, resulting in a nonsignificant difference in chi-square. Because both models fit the data equally well, the more parsimonious model (Model 3) was the preferred model.

Model 4 (paths from working memory and inhibition only) was nested within both Model 3 and Model 1. Because Model 3 was preferred over Model 1, the incremental fit of Model 4 was compared to that of Model 3. The change in chi-square between Model 3 and Model 4 was nonsignificant; therefore, Model 4 was preferred due to parsimony. Model 5 was similar to Model 4 (direct paths from working memory and inhibition to problem solving), except the correlation between working memory and inhibition was dropped. Model 5 fit the observed data significantly more poorly than Model 4, indicating that working memory and inhibition are correlated. This correlation, therefore, was retained in the remaining models.

The fit of Model 6, with a path from inhibition only and a correlation between working memory and inhibition, was compared to that of Model 4. Model 6 did not fit the observed data as well as Model 4. Similarly, Model 7, with a path from working memory to problem solving only and a correlation between working memory and inhibition, was compared to Model 4. Model 7 also did not fit the observed data as well as Model 4. Across analyses, Model 4, then, was the preferred model because it was the simplest, most parsimonious model that best fit the observed data. The estimate or coefficient (similar to a beta weight in regression analyses) for the path from working memory to problem solving was .40, and was .26 for the path from inhibition to problem solving, with a correlation of .27 between working memory and inhibition. In preschool children, this final model accounted for 29% of the variance in TOH scores.

In the multiple-group comparison, path analyses were conducted to determine whether Model 4 fit the data equally well in younger and older children. First, path coefficients (working memory to problem solving; inhibition to problem solving; correlation between working memory and inhibition) were held equal across the two age groups, resulting in an overall $\chi^2(9) = 19.27$. Then, path coefficients were allowed to differ across the two age groups, resulting in an overall $\chi^2(6) = 5.1$. The difference, $\chi^2(3) = 14.17$, between the models was significant ($p < .01$), indicating that the value of the path coefficients differed for younger and older children in Model 4 (see Figure 4). Among younger children, inhibition was the strongest predictor of problem-solving performance, with a path coefficient of .44. Working memory contributed relatively little to problem solving (path coefficient = −.05) in younger children. In contrast, among older children, working memory was the main contributor to problem solving (path coefficient = .48), and inhibition was less important (path coefficient = .07).
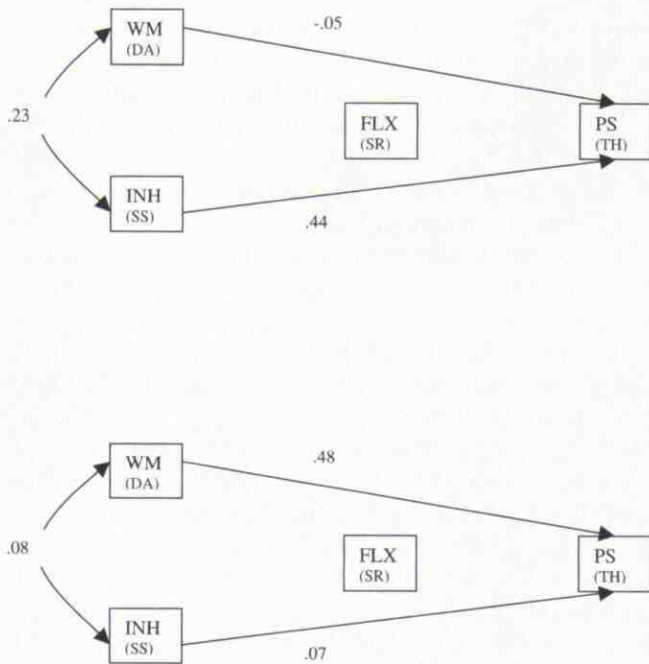
FIGURE 4   Model 8, Multiple Group Comparison Model; Top: Age ≤ 4; Bottom: Age > 4.

DISCUSSION

Path analysis was used to compare different models of executive function organization in preschool children, a population from which it can be difficult to obtain complete data on comprehensive neuropsychological batteries. Performance on a working memory task and on an inhibition task predicted performance on a complex problem-solving task, the TOH, as well as did models that included performance on a task designed to measure mental flexibility. Path analysis may be a powerful tool by which to investigate relations among executive functions or other neuropsychological abilities in populations where it is difficult to obtain complete data on a large battery of measures, such as certain clinical populations, young children, or older adults. Path analysis has many of the same capabilities as structural equation modeling, including modeling complex relations among variables, comparing complicated models to simpler, nested models, and conducting multiple-group comparisons to determine whether a model fits the observed data from different groups equally well.

Substantively regarding neuropsychology, these results provide preliminary support for those theories that are predicated on a dualistic notion of executive

functions (e.g., Diamond, 1988; Roberts & Pennington, 1996), that is, where the contributions of working memory and inhibition are central. Indeed, these constructs accounted for 29% of the variance in complex problem solving, as indicated by the TOH score, a substantial portion of performance. Furthermore, the relation between working memory and inhibition was important, consistent with Roberts and Pennington's view of a competitive balance between these two processes. However, there are other ways to conceptualize this balance, for example, with reciprocal causation, which also is consistent with some theories of prefrontal function (e.g., Fuster, 1989). Such models were not investigated here, as they can be difficult to fit, do not change the amount of variance explained in the dependent construct, and are best attempted with a full structural model that includes multiple indicators of the constructs of interest.

The benefits of directly comparing more complex models are apparent. Flexibility does not appear to mediate the relation between working memory, inhibition, and problem solving, at least in preschool children. In preschool children, mental flexibility may be less differentiated from working memory and inhibition than in school-age children and, therefore, not an important contributor to problem-solving abilities in the preschool age range. Complex problem solving and planning tasks may not demand significant mental flexibility until the child can solve problems successfully that require a larger repertoire of plans and schema (Bull et al., 2004). Alternatively, the simpler, more "motoric" flexible search behavior measured by SR (Kaufmann et al., 1989) may not adequately measure the more "cognitive" flexibility of EF. In fact, recent evidence from adult neuroimaging studies suggests that flexible switching between extradimensional categories (e.g., position to shape) activates prefrontal systems, whereas intradimensional shifts (one location vs. the other) do not (Dias, Robbins, & Roberts, 1996, 1997). However, this issue has not been studied in normal children, let alone those of preschool age. Clearly, longitudinal studies are necessary to characterize these ontogenetic relations in young children and how they map onto subsequent executive functions during the school-age years.

In contrast to a study of adults in which only inhibition contributed to TOH performance (Miyake et al., 2000), here the inhibition-only model was rejected. In preschool children, working memory contributed to complex problem solving, as did inhibition. For young children, the requirement of maintaining the task rules and their performance relative to the goal state may engage working memory to a greater extent, in comparison to the adults studied by Miyake et al. However, more complicated models of executive organization were not considered. The integration of these micro- and macrolevel constructs appears to be a fruitful avenue to explore and may yield important insights into the nature of cognitive organization and mechanisms of cognitive development.

In a related vein, the working memory-only model also was rejected. Although some theorists have emphasized the centrality of working memory in executive

control (e.g., Kimberg & Farah, 1993), such a position was not supported here, at least in this age range. This evidence must be considered preliminary given the limitations of the path model, where only one test is selected to primarily measure each construct. For example, if all of the selected tests actually measured only working memory, the observed differences between the models would simply reflect differing construct validity among the selected test instruments. Given the diverse evidence of the validity of these tasks as indicators of the constructs that make up differing aspects of executive control (e.g., Espy, 1997; Espy et al., 2001, 1999; Goldman et al., 1971; Mishkin, 1964; Miyake et al., 2000), this explanation seems unlikely.

Interestingly, the best fitting model that included inhibition and working memory as predictors of problem solving did not fit the data from younger and older children equally well. Inhibition was a stronger determinant of problem solving in younger children, whereas working memory was more important in older children. These findings may represent the different maturational timetable of these abilities. Inhibitory control may develop more rapidly in younger children, with more protracted development in working memory (e.g., Espy, 1997; Espy et al., 2001, 1999; Welsh et al., 1991). As a result, older children may bring to bear their better developed working memory to solve the more difficult, complex TOH problems, which in turn, demand more working memory for correct solution. However, these analyses were exploratory. The recommended approach is to develop and test competing models on the basis of theory or on results from an initial sample, and in the latter case, to further validate and confirm the results on a separate sample. At a minimum, a larger sample size would be required to fully test whether there are age-related differences in executive function organization.

The sample for this study included a disproportionate number of college-educated families. Although such a demographic was consistent with the local area from which children were recruited, the mean education level was higher than national expectations. It will be important to determine whether a similar pattern of executive function organization is consistent across children of varying socioeconomic circumstances. In fact, the multiple-group comparison approach demonstrated here with respect to age could be used to test the equivalency of model fit across diverse groups of children.

Although the strength of the path-analysis approach is in the ability to test structural models, there also are inherent limitations. Most importantly, performance on a single test instrument never is a pure measure of the intended construct. Although tests were chosen to primarily measure a given construct on the basis of empirical evidence and theoretical rationale, these instruments may not measure fully the hypothesized executive process. Specifically, successfully using information from a previously correct retrieval to guide performance on the next trial on DA inevitably requires a degree of inhibition, whereas successful SS Inhibit-condition perfor-

mance requires maintaining the inhibition rule online. To flexibly shift among mental sets presumably would require both maintaining the rule and inhibiting the incorrect response. Particularly with higher order executive functions, finding "pure" measures of given executive functions has proven difficult. Many of the measures used in this study have been related to the prefrontal cortex in nonhuman animals (e.g., Goldman et al., 1971; Goldman-Rakic, 1987) and in human infants (e.g., Bell & Fox, 1992), with specific performance differences noted among tasks (Espy, 1997; Espy et al., 2001, 1999) that lend a degree of convergent validity. However, exactly what cognitive functions are assessed by these measures in preschool children, and other executive function tests in older children and adults more generally, requires further explication.

Because one inherent assumption of the path model is that the selected test measures the construct of interest, if the measure is not perfectly reliable, the observed path coefficients may misspecify the true relations between constructs. Although it is possible to adjust the path models for measure reliability (i.e., by setting the paths between the construct and the measured variable to the square root of the reliability coefficient), the reliability must be known. Unfortunately, because of the experimental nature of these measures and their relatively new application to exploring cognitive organization in preschool children, reliability estimates were not available. Because a single test is used to measure the construct, vagaries related to the individual tests, such as test order or material-specific differences, may have measurable effects on model results. Although the hypothesized model accounted for a large portion of the variance in complex problem-solving abilities, the amount of residual variance also depends on test reliability, in addition to measurement and subject error, and the influences of other variables not considered in the model, for example, visuospatial ability, which has been related to TOH performance (Roennlund, Loevden, & Nilsson, 2001).

Structural equation modeling allows better assessment of the latent construct of interest by using the shared variance between multiple measures believed to assess the same cognitive construct. Recent advances in missing data techniques, such as multiple imputation (Schafer, 1997), may be useful in the application of structural equation modeling procedures to data from populations where obtaining complete data on large test batteries can be difficult, such as certain clinical populations, young children, and older adults. Some structural equation modeling programs also now include a maximum likelihood estimation routine that allows for data that are missing at random (e.g., Mplus, AMOS). Irrespective of what approach is used, the pattern of missing data must be inspected thoroughly and the assumptions that underlie such procedures considered carefully before employing such approaches. These new techniques may be a valuable method for understanding executive function organization, and cognitive processing more generally, in a wider array of populations and ages.

ACKNOWLEDGMENTS

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.

Borkowski, J. G., & Burke, J. E. (1996). Theories, models, and measurements of executive functioning: An information processing perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 235–262). Baltimore: Brookes.

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of the performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry, 45,* 743–754.

Cwik, M. (2002). *An examination of the construct validity of the Behavior Rating Inventory of Executive Function–Preschool Version.* Unpublished master's thesis, Southern Illinois University, Carbondale.

Damasio, A. R., & Anderson, S. W. (1993). The frontal lobes. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology.* New York: Oxford University Press.

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

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380,* 69–72.

Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: Restriction to novel situations and independence from "on-line" processing. *Journal of Neuroscience, 17,* 9285–9297.

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., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *Clinical Neuropsychologist, 15,* 46–58.

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.

Francis, D. J., Fletcher, J. M., & Rourke, B. P. (1988). Discriminant validity of lateral sensorimotor tests in children. *Journal of Clinical and Experimental Neuropsychology, 10,* 779–799.

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

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia, 33,* 623–642.

Golden, C. (1978). *Stroop Color Word Test.* Chicago: Stoelting.

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.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16,* 233–253.

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.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology, General, 122,* 411–428.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem solving and planning processes in preschool children. *Cognitive Psychology, 13,* 113–148.

Korkman, M., Kemp, S. L., & Kirk, U. (2001). Effects of age on neurocognitive measures of children ages 5 to 12: A cross-sectional study on 800 children from the United States. *Developmental Neuropsychology, 20,* 331–354.

Levin, H. S., Culhane, K. A., Fletcher, J. M., Mendelsohn, D. B., Lilly, M. A., Harward, H., et al. (1994). Dissociation between delayed alternation and memory after pediatric head injury: Relationship to MRI findings. *Journal of Child Neurology, 9,* 81–89.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

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

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

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wagner, 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.

Muthen, L. K., & Muthen, B. O. (1998). *Mplus user's guide.* Los Angeles: Muthen & Muthen.

Owen, A. M., Downes, J., Sahakian, B., Polkey, C., & Robbins, T. R. (1990). Planning and spatial working memory following frontal lesions in man. *Neuropsychologia, 28,* 1021–1034.

Owen, A. M., Doyon, J., Petrides, M., & Evans, A. C. (1996). Planning and spatial working memory: A positron emission tomography study in humans. *European Journal of Neuroscience, 8,* 353–364.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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. 265–282). Baltimore: Brookes.

Pennington, B. F., Bennetto, L., McAleer, O., & Roberts, R. J. (1996). Executive functions and working memory: Theoretical and measurement issues. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 327–348). Baltimore: Brookes.

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

Roennlund, M., Loevden, M., & Nilsson, L. (2001). Adult age differences in Tower of Hanoi performance: Influence from demographic and cognitive variables. *Aging Neuropsychology and Cognition, 8,* 269–283.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7,* 268–288.

Taylor, H. G., Albo, V. C., Phebus, C. K., & Sachs, B. R. (1987). Postirradiation treatment outcomes for children with acute lymphocytic leukemia: Clarification of risks. *Journal of Pediatric Psychology, 12,* 395–411.

Watanabe, M., Kodama, T., & Hikosaka, K. (1997). Increase in extracellular dopamine in primate prefrontal cortex during a working memory task. *Journal of Neurophysiology, 78,* 2795–2798.

Wechsler, D. (1987). *Wechsler Memory Scale–Revised.* San Antonio, TX: Psychological Corporation.

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. (1990). Neuropsychology of early-treated phenylketonuria: Specific executive function deficits. *Child Development, 61,* 1697–1713.

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