Neuropsychology
2008, Vol. 22, No. 4, 508–522

0894-4105/08/$12.00 DOI: 10.1037/0894-4105.22.4.508

Executive Function in Older Adults: A Structural Equation Modeling Approach

Rachel Hull
Texas A&M University

Randi C. Martin, Margaret E. Beier, and David Lane
Rice University

A. Cris Hamilton
University of Pennsylvania

Confirmatory factor analysis (CFA) and structural equation modeling (SEM) were used to study the organization of executive functions in older adults. The four primary goals were to examine (a) whether executive functions were supported by one versus multiple underlying factors, (b) which underlying skill(s) predicted performance on complex executive function tasks, (c) whether performance on analogous verbal and nonverbal tasks was supported by separable underlying skills, and (d) how patterns of performance generally compared with those of young adults. A sample of 100 older adults completed 10 tasks, each designed to engage one of three control processes: mental set shifting (Shifting), information updating or monitoring (Updating), and inhibition of prepotent responses (Inhibition). CFA identified robust Shifting and Updating factors, but the Inhibition factor failed to emerge, and there was no evidence for verbal and nonverbal factors. SEM showed that Updating was the best predictor of performance on each of the complex tasks the authors assessed (the Tower of Hanoi and the Wisconsin Card Sort). Results are discussed in terms of insight for theories of cognitive aging and executive function.

Keywords: executive function, structural equation modeling, factor analysis, dedifferentiation, cognitive aging

Do the cognitive consequences of healthy aging arise, at least in part, from a reorganization of executive functions in older adults? Executive function is conceptualized as the rather elusive set of processes that enables planning, decision-making, coordinating, sequencing, and monitoring of cognitive operations. Numerous tasks have been devised to measure these abilities, typically by requiring participants to actively maintain various rule sets, goal-states, and/or recently encountered information in the face of interfering or competing information. Several descriptive theoretical constructs have been created to describe performance outcomes that depend on these abilities, predominant among which have been *shifting* mental sets, *updating* or monitoring information, and *inhibition* of prepotent responses (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Nonetheless, researchers have continued to debate whether these executive functions represent a single underlying ability versus a set of abilities (e.g., Balota & Faust, 2001; Friedman & Mikaye, 2004; Hedden & Yoon, 2006; Miyake et al., 2000; Salthouse, 2001; Salthouse, Atkinson, & Berish, 2003), and if multiple abilities are involved, whether they are specific to different modalities (e.g., verbal vs. nonverbal, Hamilton & Martin, 2005). Other unresolved questions are whether these theoretical constructs (Shifting, Updating, and Inhibition) differentially support complex executive function tasks, and whether the recruitment of executive functions is different in older relative to younger adults (e.g., Hasher, Zacks, & May, 1999; Salthouse, 2001; West & Bowry, 2005).

To answer these questions, the present research assessed executive function in older adults on a variety of tasks. The four primary goals were to determine (a) whether tasks designed to engage individual executive functions were supported by different underlying factors, (b) which underlying skill(s) supported complex executive function, (c) whether there were systematic differences for verbal and nonverbal versions of tasks thought to tap the same specific function (e.g., inhibition), and (d) how the magnitude of relationships among tasks and the factor(s) that underlay them for older adults in this study would qualitatively compare with those previously reported in a similar study with young adults (Miyake et al., 2000).

Cognitive Aging Approaches

If one accepts that the cognitive consequences of aging may be associated with changes in executive function, the question then is

Rachel Hull, Department of Psychology, Texas A&M University; Randi C. Martin, Margaret E. Beier, and David Lane, Department of Psychology, Rice University; A. Cris Hamiton, Department of Neurology, University of Pennsylvania.

This research was supported by NIH Grant DC-00218 to R.C.M. at Rice University. The authors acknowledge the assistance of Meredith Knight, Jill Henderson, and Adam Witas in collecting data for this project, and thank Naomi Friedman and Akira Miyake for providing their experimental tasks, and Hart Blanton for insightful comments on a previous version of this paper. Portions of these data were presented at the European Conference on Cognitive Neuropsychology, Bressanone, Italy, January 2005, and at the Conference on the Place of Inhibition in Cognition at University of Texas, Arlington, March 2005.

Correspondence concerning this article should be addressed to Dr. Rachel Hull, Department of Psychology, Texas A&M University, College Station, TX 77843-4235. E-mail: rhull@tamu.edu

exactly how might executive function change with age? One idea, termed the *dedifferentiation hypothesis* (Balinsky, 1941; see also Antsey, Hofer, & Luszcz, 2003), posits that the distinction among cognitive abilities blurs with age and is reflected in significantly increased correlations between factors, or even reduction to a single factor. Salthouse and colleagues (Salthouse et al., 2003; see also Salthouse, 2001) have argued in favor of the dedifferentiation hypothesis based on their observations that performance by adults aged 18–84 years on a variety of cognitive tasks was more strongly related to fluid intelligence scores than to the executive function constructs of Inhibition, Updating, or time-sharing (dual task). However, others have argued that averaging performance in cross-sectional studies with wide age ranges can artificially inflate commonality among cognitive factors (see Anstey et al., 2003 for discussion) and can introduce age cohort confounds (e.g., Zook, Welsh, & Ewing, 2006). Moreover, a reexamination of the Salthouse et al. (2003) data found that, across age groups, executive function measures correlated better with each other than with other cognitive measures (Hedden &Yoon, 2006).

A few longitudinal studies have also investigated the dedifferentiation hypothesis, but the conclusions have been mixed. Ghisletta and Lindenberger (2003) reported dedifferentiation effects for perceptual speed and weak effects for semantic knowledge in adults age 70 – 103 years ($N = 516$). In contrast, a separate study using the same age range ($N = 1{,}823$) reported no evidence for dedifferentiation (Antsey et al., 2003) in high-ability individuals, and only inconsistent evidence in low-ability individuals (see also West & Schwarb, 2006). These outcomes suggest that cognitive ability may be a better predictor of age-related cognitive change than chronological age (Antsey et al., 2003).

Cognitive Process Approaches

Several recent studies have used an individual differences approach to augment standard experimental investigations of cognitive control. Using CFA and SEM, Miyake et al. (2000) found three partially separable factors (Shifting, Updating, and Inhibition) supporting executive function performance in young adults. The CFA showed that a three-factor model fit the data significantly better than a one- or two-factor model, or a "three independent factors" model (which sets interfactor correlations to zero). This indicated that the three factors represented distinguishable but moderately related abilities (interfactor correlations ranged from .42 – .63). For instance, repeatedly switching among different rule sets (e.g., switching between local and global figure identification in the Navon task; Navon, 1977) depended primarily on Shifting ability, whereas ignoring a very salient (but task-irrelevant) stimulus feature in order to successfully identify a less-salient but task-relevant feature (e.g., Stroop task; Stroop, 1935) depended primarily on the ability to inhibit prepotent responding to the irrelevant feature. Miyake et al. further found that the three factors contributed differentially to performance on complex cognitive tasks. Specifically, the Tower of Hanoi task (TOH, adapted from Humes, Welsh, Terzlaff, & Cookson, 1997) relied primarily on Inhibition, the Wisconsin Card Sort Task (WCST, adapted from Kimberg, D'Esposito, & Farah, 1997) depended primarily on Shifting, and the operation span task (adapted from Turner & Engle, 1989) depended most on Updating. These outcomes clearly indicated that Inhibition, Updating, and Shifting represented three partially separable factors of executive function in young adults.

Hedden and Yoon (2006) used CFA and SEM to assess the relationships between resistance to interference, cognitive abilities (i.e., verbal memory, visual memory, and processing speed), and the executive function constructs of Shifting, Updating, and Inhibition, in older adults. Their CFA revealed two executive function factors and three cognitive factors. Of the two executive function factors, one supported both Updating and Shifting abilities and the other supported Inhibition; however, their Inhibition factor supported only tasks that involved resistance to proactive interference (PI), namely, category and letter fluency tasks. In contrast, inhibition tasks that required suppression of prepotent responses (i.e., Stroop and antisaccade) correlated best with measures of perceptual speed (i.e., letter string comparisons), and thus were incorporated into a cognitive "perceptual speed" factor rather than an executive function factor. Hedden and Yoon then used SEM to determine which of the factors related to resistance to interference, which was indexed by hits and false alarms on a working memory task that incorporated PI manipulations. Hits and false alarms were significantly related to all executive function factors (Shifting/Updating and Resistance to PI), but not to all cognitive factors (hits were related to visual memory, and false alarms were related to perceptual speed). Hedden and Yoon concluded that executive function abilities could not be reduced to general cognitive abilities.

Hedden and Yoon's (2006) two-factor model of executive function (Updating/Shifting and Resistance to PI) contrasts with Miyake et al.'s (2000) three-factor model (Shifting, Updating, and Inhibition). However, the interpretation of this difference is complicated by aspects of Hedden and Yoon's experimental design. For instance, Hedden and Yoon used complex executive function tasks (WCST and the Trail Making Task) as indicators of Shifting ability, whereas Miyake et al. used only simple tasks as indicators. In general, complex tasks are impure with regard to the executive functions they tap (see Miyake et al. for discussion), and the WCST may involve different factors in older and younger adults. Thus, the inclusion of WCST and the Trail Making Test in the Hedden and Yoon study may have reduced the unique variance associated with the Shifting factor. Furthermore, the high correlation between inhibition of prepotent responses and perceptual speed (which formed the perceptual speed factor) could have resulted from variation in antisaccade performance, which requires inhibiting a reflexive eye movement in order to recognize a briefly presented target. Without comparing performance on the antisaccade task with a prosaccade version (which does not require inhibition), it is unclear whether antisaccade variation reflects differences in Inhibition ability or differences in perceptual speed. Also unclear is whether the other prepotent response inhibition task (i.e., the Stroop task) should be considered an indicator of perceptual speed. That is, the size of the Stroop effect correlates with overall reaction time, so people with slower reaction times and slower perceptual speed are also likely to show larger Stroop effects. Therefore, log transformations are necessary to eliminate the relationship between reaction time and the size of the Stroop effect in order to determine whether it is overall response time or the size of the Stroop effect that led to the loading of the Stroop task on Hedden and Yoon's perceptual speed factor.

One might argue that Hedden and Yoon's (2006) finding of a reduction of Shifting and Updating to a single factor in older adults reflects support for the dedifferentiation hypothesis. However, the strength of correlations among executive function factors (.40 – .69) was very similar to that reported by Miyake et al. (2000) for young adults (.30 – .63). This is inconsistent with the dedifferentiation position. An alternative explanation may be that, instead of devolution to a single executive function factor, age-related cognitive decline may relate to individual differences in the efficiency of working memory. Age differences in Shifting ability, which have been particularly linked to increasing global shift costs (i.e., the cost difference between task-switching blocks and single-task blocks), but not local shift costs (differences between switch and nonswitch trials within blocks with predictable switch sequences) provide evidence in favor of this view (e.g., Kray & Lindenberger, 2000; Mayr, 2001; see also Engle & Kane, 2004; Hasher et al., 1999; Kane & Engle, 2002; West, 2004). Specifically, Kray and Lindenberger suggested that the age-related increase in global shift costs results from a decreased ability to efficiently maintain competing task sets in working memory. Mayr and colleagues (Mayr, 2001; Mayr & Liebscher, 2001) elaborated that this age difference may arise because global shift costs require older adults to activate a time-consuming internal updating process to minimize interference between competing task sets, thus taxing working memory and leaving fewer resources available for keeping each task set accurate. These explanations are consistent with reports of physical changes in the frontal lobes and prefrontal dopamine systems, which are thought to trigger a decreased ability to maintain the integrity of contextual representations needed for dynamic task performance (Braver et al., 2001; see also West & Schwarb, 2006). Thus, all these researchers converge on the idea that working memory demands change with age, and this change may reflect a difference in the relative contributions of executive functions, rather than their selective elimination.

Domain-General Versus Domain-Specific Processes

Numerous cognitive models have followed from the seminal ideas of Baddeley and Hitch (1974), who proposed that a privileged "general cognitive manager" (i.e., the central executive) directs two specialized, limited-capacity subcomponents in the execution of cognitive tasks. The phonological loop specializes in maintaining verbal information (e.g., phonological, semantic, syntactic), and the visuospatial sketchpad maintains nonverbal information. Modern conceptualizations have shifted focus from passive maintenance toward active manipulation of information in working memory (e.g., Cowan, 2005), but the overarching premise that hierarchical processes control the flow of verbal and nonverbal information has persisted.

Behavioral observations from our own lab suggest that patients with semantic short-term memory deficits, which are associated with damage to the inferior frontal gyrus, may have particular difficulty inhibiting irrelevant verbal (but not nonverbal) information during executive function tasks (Hamilton & Martin, 2005; see also Martin & He, 2004, and Martin & Lesch, 1996). We have reported one such patient, ML, who showed good single-word comprehension, but deficits in the inhibition of irrelevant verbal information in the Stroop task and the recent negatives task (Hamilton & Martin, 2005). However, ML performed at normal levels in a nonverbal analogue to the Stroop task and in the antisaccade task, which required the inhibition of reflexive eye movements, suggesting there may be some distinction in executive functions that support verbal and nonverbal domains. That is, this neuropsychological dissociation suggests there are separable cognitive components that support inhibition in verbal and nonverbal domains. If so, one might expect to see some reflection of this separation in the patterns of data from neurally intact, age-matched individuals.

Indeed, neurally healthy adults have shown some evidence of a separation in verbal and nonverbal abilities, as they tend to perform about as well on simultaneous verbal and nonverbal tasks as on the same tasks performed separately; however, performance drops when both tasks are verbal or both are nonverbal (see review in Baddeley, 1998; see also Baddeley & Hitch, 1974). For example, performance on a nonverbal (i.e., nonword) task, such as tracking a moving light, would not be expected to interfere with performance on a verbal (e.g., phonological) task, such as recalling a series of letters (Baddeley, 1998). Conversely, decreased performance has typically been reported when simultaneous tasks recruit the same domain (e.g., Miyake et al., 2000; see also Smith, Jonides, & Koeppe, 1996). Fernandes, Pacurar, Moscovitch, and Grady (2006) showed that both young and older adults showed more interference for recognizing previously studied words when the concurrent task was verbal (i.e., a word animacy judgment) than when it was nonverbal (i.e., odd-digit identification). An interesting finding was no differences in the magnitude of the effect across age groups. These studies point to separable processing abilities in verbal and verbal domains, but the question to be addressed here is whether executive functions are also specialized by these domains. Some researchers have suggested that individual differences in executive function performance can be directly tied to whether the content of the test is verbal or nonverbal (see Ackerman, Beier, & Boyle, 2002). Consequently, a goal of the present research was to investigate whether analogous verbal and nonverbal executive function tasks are supported by different underlying abilities.

The Present Study

The fundamental purpose of this study was to understand the organization of executive functions in older adults, and a related goal was to assess how this organization would qualitatively compare with that reported for younger adults (Miyake et al., 2000). Therefore, we followed Miyake et al. in using only simple executive function tasks to determine the underlying factors, then separately assessing how the factors predicted performance on complex tasks (TOH and WCST). Although we used many of the same tasks as Miyake et al., a difference was that we included additional verbal or nonverbal analogues to most of the tasks in order to address our goal of determining whether any separation between verbal and nonverbal domains would be observed for healthy older adults.

Method

Participants

One hundred adults between the ages of 51 and 74 took part in this study (*M* age = 60.24 years, *SD* = 5.58). No participants had any history of neurological insults (e.g., stroke), all had normal or corrected-to-normal vision and hearing, and all were native English speakers recruited from the Houston, Texas community by announcements in two local newspapers and in bulk email solicitations to Rice University employees. In terms of educational history, 25% of participants had a high school education or its equivalent, 2% had an associate's degree, 49% had a bachelor's degree, and 24% had advanced degrees. Of the 100 participants included in the final sample, 80 were women and 20 were men.

Materials, Design, and Procedure

Participants completed 10 tasks to measure one of the three executive functions of mental set shifting (Shifting), information updating and monitoring (Updating), and inhibition of prepotent responses (Inhibition). In addition, two complex tasks commonly used to assess executive function, the WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993), and the TOH (Simon, 1975) were administered. All executive function tasks (except plus-minus) were computerized. A button box with millisecond accuracy was used to collect reaction time (RT) measures, and a voice key attached to the button box recorded RTs for verbal responses. For accuracy measures, button-press responses per task and per participant were recorded in individual computer files, and an experimenter documented on paper all spoken or pointing responses. Detailed descriptions of all the tasks are provided below. Two of the tasks (verbal keep-track and antisaccade) taken from Miyake et al. (2000) were modified, as discussed below, in order to make them easier for older adults to complete (see discussions in Chiappe, Hasher, & Siegel, 2000; May, Hasher, & Kane, 1999).

The three tasks used to identify the Shifting Factor were:

1. In the *local-global nonverbal* task (adapted from Miyake et al., 2000; see also Navon, 1977), participants saw geometric figures consisting of many small "local" figures arranged to form a much larger "global" figure (e.g., many small circles forming one large triangle). The color of the figures indicated whether participants should attend to the local (black) or the global (blue) figure. Sometimes the stimuli colors remained the same across successive trials (nonswitch trials) and sometimes they changed (switch trials). During switch trials, participants had to shift mental set from the local to the global figure or vice versa, depending on the direction of change in the color rule. Participants were to press one of two designated buttons to indicate whether the shape indicated by the color rule was a circle or a triangle, and RTs to button press were recorded by computer. A total of 309 trials were presented in a fixed random order, the first 31 trials were practice. Half the 278 target trials (i.e., trials for which responses were used in the analyses) were switch trials and half were nonswitch trials. Each incorrect response was removed, as was the response immediately following each error. Each participant's shift cost was calculated by subtracting the mean RT to nonswitch trials (no change in color rule) from the mean RT to switch trials (color rule changed).

2. The *local-global verbal* task was identical to the nonverbal version except that the figures participants saw were the letters "S" and "V." As with the nonverbal version, RTs were recorded as participants indicated the relevant figure by button press, and RTs to button press were used to calculate shift costs.

3. The *plus-minus* task (adapted from Jersild, 1927) was a paper and pencil task consisting of three lists of 30 two-digit numbers. Participants were to add 3 to each number on the first list, subtract 3 from each number of the second list, and sequentially alternate between adding 3 and subtracting 3 from the numbers on the third list. An experimenter seated behind the participant measured time on each of the three lists (in seconds) with a stopwatch. The dependent measure was shifting cost, as calculated by subtracting the mean total time on addition-only and subtraction-only lists from total time on the alternating list.

The four tasks used to identify the Updating Factor were administered as follows:

4. The *N-back nonverbal* task (adapted from Gevins & Cutillo, 1993) required participants to continuously monitor a series of tones of five different pitches and to respond when they heard any tone a second time after one intervening tone (i.e., two trials later). The task therefore required online monitoring and updating of remembered tone information. After familiarization with the five tones, participants completed 20 practice trials followed by three blocks of 20 target trials, for a total of 60 target trials. In 10 target trials, the tone heard represented a match to the tone heard two trials previously (hit trials), but did not match in the remaining 50 trials (rejection trials). Proportion of misses and false positives were separately calculated and then averaged per participant. Percent total error was the dependent measure.

5. The *N-back verbal* task was identical to the nonverbal version except that participants monitored a series of five letters (C, K, N, R, V) instead of tones, and participants completed 60 target trials, 15 of which were hit trials and 45 of which were rejection trials. Again, proportions of misses and false positives were separately calculated before averaging, and percent total error was the dependent measure.

6. The *keep-track verbal* task (adapted from Miyake et al., 2000; see also Yntema, 1963) required participants to monitor and update the last word (e.g., dog, red, Brazil, inch, platinum, mother) presented in two of six semantic categories (animals, colors, countries, distances, metals, relatives). The two relevant categories were indicated at the bottom of the screen and remained visible throughout the trial block. Participants completed two to four blocks of practice trials, followed by four blocks of 10 target trials. Prior to completing the practice and target trial blocks, participants were familiarized with each semantic category and all of the words that might appear for that category. Percent error was the dependent measure.[1]

7. The *keep-track nonverbal* task (adapted from Miyake et al., 2000; see also Yntema, 1963) was identical to the verbal task except that participants were to monitor and update the locations of two of four colors (red, yellow, green, blue) in a series of colored squares that appeared serially in a fixed random order in one of four quadrant locations on a computer screen. The two relevant colors were indicated at the bottom of the screen and remained visible throughout the trial block. At the end of each trial block, participants were to say each relevant color name and point to its last quadrant location in that trial block. Percent error was the dependent measure.

[1] In the keep track task used by Miyake et al., (2000), participants were required to monitor for three, four, and five categories. However, pilot data showed that many older participants performed near floor with this many categories, and thus the number of categories was reduced to two for the present study.

The three tasks used to identify the Inhibition Factor were administered as follows:

8. The *Stroop nonverbal* task (Hamilton & Martin, 2005) was adapted from the original Stroop (1935) task, and similar nonverbal adaptations of this task have been used by Clark and Brownwell (1975) and Lu and Proctor (1994). In this task, participants saw a series of directional arrows. The arrows pointed either left or right, and they appeared either on the left side, the right side, or in the center of the screen. Thus, arrows could either be congruent with spatial location (left-pointing arrow on left side of screen) or incongruent (left-pointing arrow on right side of screen) or neutral (left-pointing arrow in center of screen). As quickly as possible, participants were to press a key corresponding to the spatial location of the arrow on the screen, regardless of the direction in which the arrow was pointing. Responses and RTs in milliseconds were recorded into a computer file. After 14 practice trials, participants completed three blocks of 81 target trials for a total of 243 target trials. There were 80 congruent trials, 83 incongruent trials, and 80 neutral trials. Incorrect responses were removed, and the dependent measure was the difference in RT between correct responses to incongruent and neutral trials.

9. In the *Stroop verbal* task (adapted from Miyake et al., 2000; see also Stroop, 1935), participants saw a series of words (RED, GREEN, BLUE, PURPLE, ORANGE, or YELLOW) and asterisk strings in the center of the screen. Each word or asterisk string appeared in one of the six aforementioned font colors; thus, stimuli could either be congruent with font color (RED in red font) or incongruent (RED in blue font) or neutral (asterisks in any color font). Participants were to name the font color as quickly as possible, and ignore the written word. RTs in milliseconds were recorded by voice key, and an experimenter recorded verbal responses for accuracy. After 14 trials for practice and voice key calibration, participants completed three blocks of 81 target trials for a total of 243 target trials. There were 12 congruent trials, 65 incongruent trials, and 39 neutral trials. Incorrect responses were removed, and the dependent measure was the difference in RT between correct responses to incongruent and neutral trials.

10. In the *antisaccade* task (adapted from Miyake et al., 2000; see also Roberts, Hager, & Heron, 1994), participants saw a central fixation point followed by an attention-capturing visual cue (black square) displayed for 175 ms on either the left or right side of the screen. Participants were to immediately move their eyes and attention to the opposite side of the screen in order to detect an arrow stimulus displayed there for 175 ms, after which it was masked by a gray square.[2] The short duration of the arrow stimulus required that participants inhibit the reflexive response of focusing on the initial visual cue in order to detect the arrow stimulus. The arrow stimulus pointed either up, left, or right, and participants were to press a key corresponding to the direction of the arrow. Head movements were reduced using a chinrest. Responses and RTs in milliseconds were recorded into a computer file. After 10 practice trials, participants completed three blocks of 42 target trials for a total of 126 target trials. Proportion of errors was the dependent measure.

To clarify that performance on the antisaccade task reflected the ability to inhibit prepotent responses (reflexive eye movements), rather than merely the ability to recognize a briefly presented stimulus, performance data for the *prosaccade* task were collected from a subset of 36 participants. The prosaccade task was identical to the antisaccade task except that the visual cue and the arrow stimulus appeared on the same side of the screen. Thus, there was no need to inhibit looking toward the initial visual cue; in fact, looking in that direction would improve arrow stimulus detection. As with the antisaccade task, participants completed 10 practice trials and then three blocks of 42 target trials for a total of 126 target trials, and proportion of errors was the dependent measure.

Complex Tasks

In the *Wisconsin Card Sorting* task (adapted from Miyake et al., 2000; see also Heaton et al., 1993), participants were to sort cards into categorized piles according to different sorting rules. Four reference cards containing variously colored shapes were displayed at the top of the screen throughout the task and represented the four piles into which test cards could be sorted. The sorting rules were color (red, green, blue, or yellow), shape (circle, square, star, or cross), or number (1, 2, 3, or 4), and only one sorting rule was relevant at a time. On each trial (including 30 practice trials), a test card was presented in the middle of the screen below the reference cards. Participants were to sort the test card into the appropriate pile by using the mouse to drag and drop the test card onto the reference card that met the current sorting rule. Participants determined the current sorting rule by dropping test cards onto different piles and then receiving feedback ("RIGHT" or "WRONG" appeared below the chosen pile) until the correct rule was discovered. After eight correct sorts were consecutively completed, the sorting rule changed without warning, although participants were not explicitly informed of this criterion. The process was repeated until 15 sorting rules were successfully completed or total trials exceeded 288. Participants were given as much time as needed to complete the task. The dependent measure was the number of perseverative errors—that is, continued use of a given sorting rule after feedback indicated it was no longer the current rule.

In the *Tower of Hanoi* task (adapted from Miyake et al., 2000; see also Simon, 1975), participants were to move a set of disks from one of three pegs (initial state) to another (goal state) according to specific movement rules. First, participants saw a picture of the initial and goal states, and then they saw the initial state on the computer screen. The disks were graduated in size and stacked like a pyramid on the first peg. The movement rules were that only the top disk on any peg could be moved, only one disk at a time could be moved, and no disk could ever be placed on top of a smaller disk, thus requiring participants to make some counterintuitive moves in order to reach the goal state (e.g., initially putting the small disk on the goal peg even though the goal state required it to be at the top of the stack). The experimenter demonstrated disallowed moves (e.g., putting a larger disk onto a smaller disk), and how to use the mouse to perform allowed moves (e.g., moving the first disk onto the third peg), after which participants were given two practice sessions. This was followed by three target sessions, involving three, four, and five disks, respectively. Participants

[2] The corresponding times used by Miyake et al. (2000) were 150 ms for the cue and the stimulus duration. The times were lengthened in the present study because pilot data showed low accuracy for older participants with the shorter times.

were given as much time as needed to complete the task. The number of moves to achieve the goal state was the dependent measure.

Verbal and nonverbal knowledge. To estimate the degree to which individual differences in verbal and nonverbal knowledge might influence performance on verbal and nonverbal versions of executive function tasks, and to assess whether knowledge in general would correlate with particular aspects of executive function ability (as has been shown for young adults; see Friedman, Miyake, Corley, Young, DeFries, & Hewitt, 2006), two WAIS-III subtests (Weschler, 1997) were administered to a subset of participants ($N = 46$). The vocabulary subtest (VOC) is an expressive measure of verbal intelligence, where performance on defining each of 33 words was scored as 0, 1, or 2 for a maximum of 66 points. The picture completion subtest (PC) is an expressive measure of knowledge, where the ability to identify one vital, missing detail in each of 25 pictures was scored as 0 or 1 for a maximum of 25 points. According to Johnson and Bouchard (2005), PC taps an intellectual factor related to visualization ability. Both tests were administered according to the standard instructions provided in the WAIS-III manual (Weschler, 1997).

General Procedure

Participants were tested individually in three sessions and were paid $7.50 per hour. Each session lasted approximately 2 hrs for a total of 6 hrs of testing. Participants were seated at a table in front of a computer screen, and an experimenter was seated to one side. To minimize any measurement error resulting from participant-by-order interactions, all tasks were administered in the same order, and tasks meant to measure the same executive function were never presented consecutively. In the first session, participants completed verbal *N*-back, nonverbal local-global, and then WCST. In the second session they completed nonverbal Stroop, TOH, verbal keep-track, and then antisaccade. In the third session they completed nonverbal *N*-back, verbal local-global, nonverbal keep-track, verbal Stroop, and then plus-minus. Finally, for participants able to return for a fourth session, the VOC, PC, and prosaccade tasks were administered in order to serve as validation measures to test for any influence of verbal ($N = 46$) or nonverbal ($N = 46$) knowledge, and the ability to carry out visual saccades and detect briefly presented targets ($N = 36$), respectively.

Interpolation of Missing Data

In some cases, data for individual participants were incomplete for one or more of the tasks. For the verbal and nonverbal local-global, verbal Stroop, and antisaccade tasks, some observations were lost because of equipment malfunction. In the few remaining cases, participants were unable to return for testing because of personal conflicts. Thus, it was not the case that missing data were the result of an inability to execute the given task. However, for the three validation tasks (VOC, PC, prosaccade), less than half the sample participated. Therefore, for the tasks presented to all participants (i.e., all but the validation measures) we computed separate group means for the group that completed the validation measures and the group that did not. An analysis of variance (ANOVA) showed no group differences, $F(1, 98) = 0.10$, $p = .75$. Therefore, we believe it was reasonable to predict missing scores for the tasks included in the CFA on the basis of an individual's observed scores on other tasks (scores were not imputed for the validation tasks). The full information maximum likelihood method (FIML) was used for the CFA and SEM analyses to derive missing values that represented scores that were minimally deviant from the observed scores for a given individual.[3] This procedure provided less than 15% of scores.

Transformations and Outlier Analyses

To achieve normal distributions of observations necessary for the multivariate techniques used in this study, transformations on data from each variable were conducted as follows. For correct RT responses, we first computed the mean RTs for each participant in each condition of each variable. Outliers were then identified by participant and by condition, where values less than or greater than 2.5 *SD*s from the condition mean for that participant were removed. The rationale for removing rather than replacing outliers was that, at the level of the individual, outliers indicated errors of anticipation (extremely short RTs) or loss of task goals (extremely long RTs), making such observations inappropriate for inclusion in the analyses. This trimming procedure allowed us to retain important individual differences between participants while minimizing the effects of extreme observations within participants. This procedure affected less than 2% of RT observations. Finally, log transformations were used to achieve normal distributions for the verbal and nonverbal keep-track tasks, the verbal and nonverbal *N*-back tasks, the verbal and nonverbal Stroop tasks, and the TOH and the WCST.

Results

The goals of our analyses were to examine whether the performance of older adults on different executive function tasks reflected a single underlying ability or a set of (semi) independent abilities, to evaluate how any underlying factors would contribute to complex task performance, to assess whether the pattern of results would be similar to that obtained with younger adults (Miyake et al., 2000), and to determine whether verbal and nonverbal versions of executive function tasks were supported by the same or distinct underlying abilities (Hamilton & Martin, 2005). To address these goals, our results are described in three sections. First, descriptive statistics, reliability estimates, and correlations among all variables are presented and discussed. Next, a CFA of the simple executive function measures is presented. Finally, a SEM analysis is discussed, including examination of the factors identified in the CFA as predictors of performance on two different complex executive function tasks (i.e., WCST and TOH).

Preliminary Data Analysis and Task Correlations

Means and standard deviations for the simple and complex measures of executive function and for VOC and PC are presented

[3] Because the FIML method does not return the values it estimates in observable form, we also employed the estimation maximization (EM) method to obtain the actual imputed values needed to produce the correlation matrix reported in Table 2. The justifications for this decision were: (a) the matrix is necessary to allow readers to replicate our results or test competing models, and (b) in LISREL, FIML uses EM to derive starting values.

in Table 1 (reliability estimates for the simple measures are also provided). We did not collect item level data for the TOH, WCST, VOC, and PC measures and therefore do not report reliabilities for these measures. However, these measures have been used extensively in prior research and show reasonable validity and reliability in assessing ability (Rvan & Ward, 1999; Salthouse et al., 2003; Weschler, 1997) and complex executive function (Brennan, Welsh, & Fisher, 1997; Miyake et al., 2000; Ozonoff, 1995).

Correlations among the simple and complex executive tasks and validation measures are shown in Table 2. Because scores for the VOC and PC validation measures were calculated in terms of number correct, negative correlations indicated better performance on executive function measures, all of which were calculated in terms of error scores or RT cost. There are several findings of particular note in Table 2. First, consistent with previous reports in the literature (e.g., Miyake et al., 2000), intertask correlations were generally low between tasks thought to load on different factors (median −.06) but were considerably higher among tasks within a single factor (median .40), which suggests that the measures used in this study were valid indicators of the underlying factors they were presumed to engage. One exception to this, however, was the plus-minus task, which did not have significant correlations with other measures selected to identify the Shifting factor (i.e., correlations of .07 and −.10 with local-global verbal and local-global nonverbal), but rather had the highest correlations with measures selected to identify the Updating factor. This pattern of correlations indicated that the plus-minus task would not load significantly on the Shifting factor, but would instead load on the Updating factor. Therefore, plus-minus was treated as an Updating task in subsequent analyses. Although this represents a shift from a confirmatory to an exploratory approach, we feel it was a logical choice given the patterns of correlations associated with this measure. Moreover, we believe this is theoretically plausible because Miyake et al. (2000) also found significant relationships between plus-minus and all three of their Updating tasks, and because plus-minus is similar to Updating measures in terms of the requirement to keep task sets/representations from previous trials active during the current trial. Specifically, information from previous trials is needed to select the appropriate task set for the current trial in plus-minus, and to compare to the current trial in keep-track and *N*-back tasks. In contrast, it is not necessary to keep previous trial information active in order to complete any of the Shifting tasks (we return to this issue in more detail in the General Discussion).

Second, although the verbal keep-track task (an Updating indicator) had been made substantially easier than the corresponding task used in Miyake et al. (2000), there was no striking change in the pattern of correlations with other variables. Specifically, the keep-track task in Miyake et al. correlated with one other Updating task (letter memory) and with one Shifting task (plus-minus), but it did not correlate significantly with any of the Inhibition tasks. The revised keep-track (verbal) task used here correlated significantly with all of the other Updating tasks and one Inhibition task, and had a significant negative correlation with one Shifting task

Table 1
Descriptive Statistics for the Executive Function Tasks and Validation Measures

Variable	Mean	*SD*	Range	Reliability
		Simple executive function tasks		
Local-global verbal (ms)	202.22	167.20	−183.89, 668.02	.78
Local-global nonverbal (ms)	217.12	166.05	−130.65, 569.58	.55
Plus-minus (sec)	41.02	10.07	26.50, 73.50	.95[d]
Keep-track verbal	6.55	5.67	0.00, 21.25	.78[a]
Keep-track nonverbal	11.28	8.41	0.00, 48.75	.84[a]
N-back verbal	13.70	9.37	3.33, 44.44	.71
N-back nonverbal	24.11	12.82	2.00, 59.00	.74
Stroop verbal (ms)	243.68	102.69	56.36, 491.61	.90
Stroop nonverbal (ms)	63.22	60.91	−100.05, 260.01	.98
Antisaccade	0.28	0.15	0.02, 0.67	.90
		Complex executive function tasks		
Tower of Hanoi	87.67	24.48	54.00, 163.00	—
Wisconsin Card Sort Task	6.56	4.66	0.00, 21.00	—
		Validation measures		
Prosaccade[b]	0.04	0.02	0.02, 0.09	.96
Vocabulary[c]	52.28	7.87	31.00, 65.00	—
Picture Completion[c]	21.70	2.85	13.00, 25.00	—

Note. $N = 84 – 100$ depending on missing data unless otherwise noted.
[a] $N = 60$ for reliability estimates only, as item level data were not available for the entire sample. [b] $N = 36$. [c] $N = 46$. The data analyses used trimmed RTs for local-global, plus-minus, and Stroop tasks. Proportion error was used for antisaccade and prosaccade. Percent error was used for keep-track and *N*-back tasks. Number of perseveration errors was used for WCST. Number correct was used for vocabulary and picture completion. Number of moves was used for Tower of Hanoi. Reliability estimates calculated by using split half reliability methods with Spearman Brown Prophecy formula applied. [d] Reliability estimate obtained by calculating an average correlation between three test parts (via r to z transformation) and applying the Spearman-Brown Prophecy formula.

Table 2

Correlations for Simple and Complex Executive Function Tasks, Validation Measures, and Age

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
Shifting factor																
1. Local-global V	—															
2. Local-global NV	.49**	—														
Updating factor																
3. Plus-minus	.07	−.10	—													
4. Keep-track V	−.22*	−.10	.45**	—												
5. Keep-track NV	−.14	.06	.30**	.42**	—											
6. *N*-back V	−.18	.05	.06	.40**	.45**	—										
7. *N*-back NV	−.05	−.05	.17	.40**	.21*	.40**	—									
Inhibition factor																
8. Stroop V	−.14	−.11	.24*	.24*	.16	.20*	−.03	—								
9. Stroop NV	−.01	.02	.05	−.06	.19	.16	−.12	.27**	—							
10. Antisaccade	.16	.26*	.02	.18	.15	.26**	−.22*	.13	.09	—						
Complex executive function tasks																
11. TOH	−.34**	−.05	−.03	.43**	.25*	.45**	.22*	.20*	.02	−.01	—					
12. WCST	−.19	.06	−.01	.36**	.36**	.39**	.19	.30**	.00	.24*	.43**	—				
Validation tasks																
13. Prosaccade[a]	−.29**	−.25*	.09	.32**	.58*	.33**	−.01	.14	.03	.53**	.01	.47**	—			
14. Vocabulary[b]	−.04	−.15	−.27**	−.51**	−.42*	−.56**	−.35**	−.13	.03	−.22*	−.20*	−.37**	−.32**	—		
15. Picture completion[b]	−.07	−.05	−.23*	−.62**	−.52**	−.29**	−.15	−.31**	−.11	−.22*	−.34**	−.29**	−.32**	.36	—	
Age																
16. Age	−.02	.08	.22*	.01	.05	−.02	−.24*	.10	.14	.53**	−.18	.04	.37**	-.28**	−.00	—

Note. WCST = Wisconsin Card Sort Task; TOH = Tower of Hanoi; V = verbal; NV = nonverbal.
[a] $N = 36$. [b] $N = 46$.
* $p < .05$; ** $p < .01$, $N = 100$.

(see General Discussion for more on possible reasons for a negative relationship between Shifting and Updating tasks in older adults). However, greater deviation in the pattern of correlations was evident for the other task that had been modified—that is, the antisaccade task—even though the change in the task was minor (i.e., the addition of 25 ms to the cue and target durations). The antisaccade task in Miyake et al. (2000) correlated significantly with the two other Inhibition tasks, although the size of these correlations was modest (i.e., .20 and .22). It also had correlations of similar size with two of the Updating tasks (i.e., .26 and .22). For our antisaccade task, we observed nonsignificant correlations with the two other Inhibition tasks, significant positive correlations with one of the Shifting tasks and one of the Updating tasks, and a significant negative correlation with another Updating task. It seems unlikely that the source of this change was the minor modification to the antisaccade task. As will be discussed further below, there is generally less clear evidence of an Inhibition factor in the data from these older participants (see also Hedden & Yoon, 2006).

Third, there was no evidence for underlying verbal and nonverbal factors for the measures. That is, across simple executive function tasks, the mean correlation among verbal measures (calculated via *r*-to-*z* transformation) was .00 and the mean correlation among nonverbal measures was −.01, In contrast, the mean correlation *between* verbal and nonverbal measures was .21. Thus, on average, measures that used verbal content did not correlate more highly with other verbal measures than with nonverbal measures, and vice versa, as would be expected if executive functions were supported by separable verbal and nonverbal factors.

A fourth interesting outcome in Table 2 concerns the tasks we have termed "validation measures" (i.e., prosaccade, VOC, and PC). These three measures were administered to a subset of participants in order to serve as additional checks that our antisaccade task was functioning as intended and that our Updating tasks were associated with intelligence measures, consistent with other reports in the literature. The prosaccade task was included to test whether variation on the antisaccade task (presumed to measure the ability to inhibit prepotent responses, i.e., reflexive eye movements) might instead reflect merely a decreased ability to quickly perceive stimuli as a consequence of general aging. Our older participants showed high accuracy on the prosaccade task ($M = 98\%$) in line with previous outcomes with young adults ($M = 97\%$; Hamilton & Martin, 2005), indicating that the older individuals in our sample did not demonstrate an age-related slowing in the recognition of rapidly presented stimuli. This outcome provides reasonable confidence that our antisaccade measure reflected inhibition of a prepotent response rather than simply an ability to detect a briefly presented target.

The VOC and PC measures were used to estimate the degree to which individual differences in verbal and nonverbal ability might influence performance on verbal and nonverbal versions of executive function tasks, and to check that our Updating tasks were significantly associated with verbal and nonverbal knowledge in our older adult sample, as has been reported for young adults (e.g., Friedman et al., 2006, found that Updating ability was highly correlated with intelligence measures in young adults, whereas Shifting and Inhibition were not). It is important that the indicator tasks for Updating in the Friedman et al. study (i.e., keep-track, *N*-back, and letter memory) were similar to those used in the present study. We used VOC and PC from the WAIS-III for our respective measures of verbal and nonverbal knowledge because we wanted to examine whether the simple executive function tasks used in this study had common variance in terms of verbal and nonverbal content.

Our outcomes revealed a robust relationship between intelligence and Updating skill in older adults (see Table 2). Specifically, the correlations between the VOC and PC measures and each of the Updating measures were significant and generally of medium to large magnitude (median .39; Cohen, 1992), with the single exception that nonverbal *N*-back was not related to PC. In contrast, there was no systematic relationship between Shifting or Inhibition indicators and the two intelligence measures. These outcomes are consistent with the findings of Friedman et al. (2006) with young adults, suggesting that Updating skills may be particularly associated with intelligence for young and old adults alike.

It is interesting that the verbal and nonverbal versions of our Updating tasks did not show differential relationships with our verbal and nonverbal intelligence measures. We offer two possible explanations. First, it may be that one or both of the intelligence tasks (VOC and PC) did not adequately measure any differential underlying abilities that may have contributed to our verbal and nonverbal Updating tasks. However, in light of the lack of cohesiveness among the verbal and among the nonverbal measures (as described in point three above), we think a more likely possibility is that the lack of differentiation here reflects the absence of evidence for separable verbal and nonverbal factors in our sample.

A final point of interest from Table 2 concerns the lack of relationship between age and executive function performance in our sample (age was correlated with only 3 of the 10 indicator tasks, and none of the complex tasks). However, we caution that any age effects could have been obscured by the restriction of the age range in our sample (i.e., only adults over age 51 and an *SD* of only 5.58). Thus, these results cannot rule out a relationship between age and executive function performance in general.

Confirmatory Factor Analysis

A CFA was conducted using LISREL 8.7 (Jöreskog & Sörbom, 2001) to test the fit of a three factor model (Shifting, Updating, and Inhibition) with 10 indicators loading on the factors as follows: (a) Shifting—local-global verbal and local-global nonverbal; (b) Updating—plus-minus, keep-track verbal, keep-track nonverbal, *N*-back verbal, and *N*-back nonverbal; and (c) Inhibition—verbal Stroop, nonverbal Stroop, and antisaccade. The FIML technique for handling missing data was used because this method has been shown to provide better parameter and standard error estimates than other methods in SEM (Newman, 2003). Because the number of participants who completed the prosaccade ($N = 36$), VOC ($N = 46$), and PC ($N = 46$) validation measures was quite limited, and because these were not measures of primary interest in determining the factor structure of executive function in older adults, these were excluded from the CFA analysis. The initial test of the model resulted in a nonsignificant negative error variance associated with the LGV indicator (i.e., an inadmissible solution). Because the negative error variance was not significantly different from zero, and because we could reasonably estimate this error variance based the reliability of the local-global verbal measure, we fixed the error term associated with local-global verbal to $(1 - r_{xx'}) * \sigma^2$ where σ^2 is the variance associated with the measure (Bollen, 1989). The resulting model is shown in Figure 1. As can be seen in the figure, all paths associated with indicators of the Shifting and Updating factors were significant. However, the path for only one indicator of the Inhibition factor (verbal Stroop) was significant. This is not surprising given the low to moderate correlations between these indicators (verbal Stroop, nonverbal Stroop, and antisaccade, see Table 2). Fit of the model was adequate $\chi^2(N = 100, 33) = 49.29$, $p = .03$, Root Mean Squared

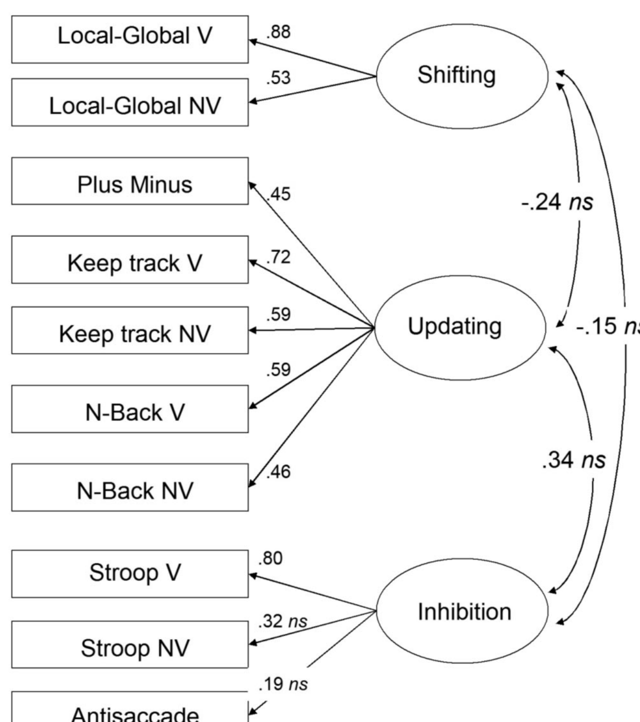

Figure 1. Confirmatory factor analysis of executive function tasks. All paths are significant except where noted. *ns* = not significant; V = verbal; NV = nonverbal.

Error of Approximation (RMSEA) = .07.[4] Following Byrne (1998), nonsignificant chi-square values indicate that there was no significant difference between the sample covariance matrix and the restricted covariance matrix as specified in the model and represent a good fit of the data to the model. RMSEA values between .00 and .05 represent good model fit and those over .05 to .08 represent adequate model fit (MacCallum & Austin, 2000). Because samples of 100 or fewer can lead to wide confidence intervals and reduced precision, Table 3 provides a detailed description of the path coefficients with 95% confidence intervals for the CFA outcomes to clarify the precision and accuracy of our path estimates and associated interpretations.

In order to ensure that the three-factor solution provided the best fit for the data, we compared the three-factor CFA model to reduced models with one and two factors. We first compared our three-factor model shown in Figure 1 with a model having only one general factor. Fit of a model with only one general factor was poor, $\chi^2(N = 100, 36) = 119.01$, $p > .05$, RMSEA = .15. A chi-square difference test confirmed that the model with three factors fit the data significantly better than this one-factor solution, $\Delta\chi^2(N = 100, 3) = 69.72$, $p < .001$. Following Miyake et al. (2000), we also compared the fit of the three-factor model shown in Figure 1 with all of the different possible variations of a two-factor model including, (a) a model with a combined Updating and Inhibition factor and a separate Shifting factor (Model A); (b) a model with a combined Shifting and Inhibition factor and a separate Updating factor (Model B); and (c) a model with a combined Shifting and Updating factor and a separate Inhibition factor (Model C). Evaluation of the change in χ^2 statistics for each of these models versus the three-factor model shown in Figure 1 revealed that the three-factor solution fit the data significantly better than two of the two-factor solutions, Models B and C. Change in χ^2 for these comparisons were $\Delta\chi^2(N = 100, 2) = 10.21$, $p < .01$ for Model B, $\Delta\chi^2(N = 100, 2) = 62.88$, $p < .01$ for Model C. The fit of Model A was not significantly different from the three-factor solution, $\Delta\chi^2(N = 100, 2) = 4.89$, $p > .05$. However, in Model A, as in the three-factor solution, the only indicator of the Inhibition factor with a significant loading on the combined factor (Updating and Inhibition) was verbal Stroop. Although significant, the path coefficient of this loading was small (i.e., .28). Because of this small path coefficient and the failure of the other indicators of the Inhibition factor to have significant loadings, we determined that the Inhibition factor was not adequately determined in this study either because our measures were insufficient or due to issues relating to the sample. Thus, we did not include this factor in further analyses. Again, we acknowledge that this represents a shift from a confirmatory to an exploratory approach, but we feel that the shift makes sense from a theoretical standpoint that will be further discussed in the General Discussion section.

Table 3
Confidence Intervals Surrounding Path Coefficients in the Confirmatory Factor Analysis of Executive Function Tasks

		Confidence intervals	
	Path coefficient	Upper bound	Lower bound
Indicator to factor			
Local-global V to Shifting	.88	1.06	0.70
Local-global NV to Shifting	.53	0.75	0.31
Plus-minus to Updating	.45	0.72	0.18
Keep-track V to Updating	.72	0.94	0.50
Keep-track NV to Updating	.59	0.81	0.37
N-back V to Updating	.59	0.81	0.37
N-back NV to Updating	.46	0.70	0.22
Stroop V to Inhibition	.80	1.53	0.07
Stroop NV to Inhibition	.32 *ns*	0.67	−0.03
Antisaccade to Inhibition	.19 *ns*	0.52	−0.14
Factor to factor			
Shifting to Updating	−.24 *ns*	0.01	−0.49
Shifting to Inhibition	−.15 *ns*	0.16	−0.46
Updating to Inhibition	.34 *ns*	0.71	−0.03

Note. All paths are significant except where noted. *ns* = not significant; V = verbal; NV = non-verbal.

After eliminating the Inhibition factor, the remaining two-factor CFA solution with Shifting and Updating and the same seven indicators used for the three-factor solution fit the data well, $\chi^2(N = 100, 14) = 17.76$, $p > .05$, RMSEA = .05. All path coefficients from the indicators to the factors in this model were significant. Although the fit of the model was good, there may be a concern that the factors were determined mainly by shared method variance between similar indicators (e.g., the local-global, keep-track, and *N*-back tasks). For the local-global task, it was not possible to examine the impact of method variance because the only two indicators of the Shifting factor were the local-global tasks. To examine the effect of method variance on the Updating factor, we correlated the residuals of similar tasks (e.g., keep-track, *N*-back). These correlated residuals did not render the path coefficients from these indicators to the Updating factor nonsignificant. Furthermore, for the keep-track tasks, there was no evidence of significant common method variance, perhaps a reflection of the difference in these two tasks (one requiring processing of verbal categories, and the other requiring processing of spatial information). Thus, we concluded that although there is some common method variance associated with the *N*-Back tasks, it is not enough to affect the Updating factor in any substantive way.

The correlation between the Shifting and Updating factors in the two factor model was small and negative (−.24), but not significant. This nonsignificant relation is evidence of minimal overlap between these factors, a finding that is in direct contrast with the dedifferentiation hypothesis, which predicts large interfactor correlations in older adults.

Structural Equation Model

To examine the relations between the Shifting and Updating factors and the complex executive function tasks, we conducted SEM analysis using LISREL. This approach mirrored that used by Miyake et al. (2000) specifically to allow us to compare the trends of performance between the present sample of older, community dwelling adults to those Miyake et al. obtained with a young adult

[4] Relative fit indices such as CFI, NNFI, NFI are inappropriate when FIML is used as the estimation method in SEM because FIML estimation violates assumptions needed to calculate these indices.

college student sample.[5] Because we were interested in how the simple executive function tasks predicted different complex tasks, but were not necessarily interested in the relation between the two complex tasks, we ran two separate models: one with TOH as the criterion and another with WCST. The models and path coefficients are shown in Figure 2. Values on top are representative of the model with TOH as the criterion; values on the bottom represent the model with WCST as the criterion. Fit of both models was good: TOH, $\chi^2(N = 100, 19) = 27.45$, $p > .05$; RMSEA = .068; WCST, $\chi^2(N = 100, 19) = .08$, $p > .01$; RMSEA = .079. Because of the relatively limited number of participants in our sample, we have provided a detailed accounting of the path coefficients and 95% confidence intervals for the models with TOH as the criterion (see Table 4) and WCST as the criterion (see Table 5) to clarify that our estimates were reasonably precise.

As shown in Figure 2, both TOH and WCST were best predicted by the Updating factor, which had a significant path coefficient in both models. Although the path from Shifting to the complex tasks was larger for TOH than for WCST, Shifting was not significantly related to performance for either the TOH or WCST task. This is in contrast to Miyake et al. (2000) who found that Shifting was the most important predictor of performance on the WCST and Inhibition was the most important factor for TOH performance.

General Discussion

The overarching purpose of the present study is to advance understanding of the organization of executive functions in older adults. In particular, one goal of this research was to investigate whether different executive functions were better explained by a single-factor solution versus by multiple underlying factors. A second goal was to determine which underlying skill(s) contributed to the performance of complex executive function tasks. A third goal was to evaluate whether performance would differ reliably between simple tasks designed to target the verbal versus the nonverbal domains. A final important goal of this research was to investigate whether findings with older adults would replicate those with younger adults tested on similar tasks (Miyake et al., 2000).

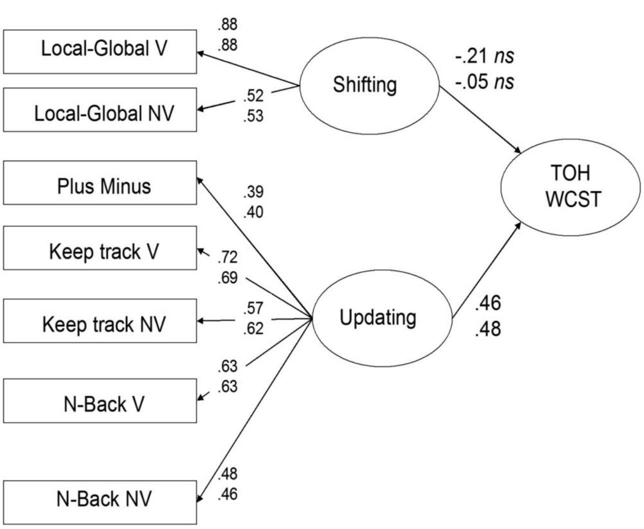

Figure 2. Structural equation models with two executive function factors predicting performance on the Tower of Hanoi (TOH) and Wisconsin Card Sort Task (WCST). Values above are those associated with the TOH task; values below are associated with the WCST. All paths are significant except where noted. *ns* = not significant; V = Verbal, NV = nonverbal.

Table 4
Confidence Intervals Surrounding Path Coefficients in the Structural Equation Model of Predicting TOH Performance

	Path coefficient	Confidence intervals: Upper bound	Confidence intervals: Lower bound
Indicator to factor			
Local-global V to Shifting	.88	1.06	0.70
Local-global NV to Shifting	.52	0.74	0.30
Plus-minus to Updating	.39	0.66	0.12
Keep-track V to Updating	.72	0.94	0.50
Keep-track NV to Updating	.57	0.79	0.35
N-back V to Updating	.63	0.85	0.41
N-back NV to Updating	.48	0.70	0.26
Factor to TOH			
Shifting to TOH	−.21 *ns*	0.01	−0.43
Updating to TOH	.46	0.70	0.22

Note. All paths are significant except where noted. *ns* = not significant; V = verbal; NV = non-verbal; TOH = Tower of Hanoi.

Executive Function Factor Structures in Older (and Younger) Adults

The present results indicate that different kinds of simple executive function tasks are differentially supported in older adults by two largely separable factors (i.e., Updating and Shifting). More specifically, the Shifting factor is most related to the ability to use external cues to activate one of two competing rule sets in the local-global tasks, as measured by local shift costs. The Updating factor is most related to the ability to maintain and update information in working memory, as measured by the *N*-back and keep-track tasks, and the plus-minus task, which measures global shift costs. Thus, the answer to our first specific goal is that a single factor solution is insufficient to explain executive function performance, as was also found by Miyake et al. (2000). However, our pattern of results marks the conspicuous absence of a third factor that was detected for young adults, namely, Inhibition.

An important question is why the Inhibition factor did not emerge in our study with older adults. There are at least two possibilities. First, the lack of cohesiveness among our Inhibition measures (i.e., verbal and nonverbal Stroop and antisaccade), and the consequential failure of these to indicate a latent Inhibition variable, may reflect a deficit in inhibition in older adults. However, that would necessitate not only a lower level of performance

[5] It is important to note that, relative to the Miyake et al. (2000) study, differences in terms of the measures (i.e., we added verbal or nonverbal analogues to each of the latent variable indicators and eliminated other indicators for which analogues were not feasible) and samples (i.e., sample variances are not equivalent to population variances) disallow direct statistical comparisons between the two studies.

Table 5
Confidence Intervals Surrounding Path Coefficients in the Structural Equation Model of Predicting WCST Performance

	Path coefficient	Confidence intervals Upper bound	Confidence intervals Lower bound
Indicator to factor			
Local-global V to Shifting	.88	1.06	0.70
Local-global NV to Shifting	.53	0.75	0.31
Plus-minus to Updating	.40	0.67	0.13
Keep-track V to Updating	.69	0.91	0.47
Keep-track NV to Updating	.62	0.84	0.40
N-back V to Updating	.63	0.85	0.41
N-back NV to Updating	.46	0.68	0.24
Factor to WCST			
Shifting to WCST	−.05 *ns*	0.19	−0.29
Updating to WCST	.48	0.72	0.24

Note. All paths are significant except where noted. *ns* = not significant; V = verbal; NV = non-verbal; WCST = Wisconsin Card Sort Task.

on inhibition tasks, but also less variance in its indicators relative to those obtained by Miyake et al. (2000). This did not appear to be the case for the two inhibition measures that overlapped across studies (i.e., Stroop and antisaccade). The standard deviation for the raw score Stroop effect was 102 ms in the present study and 60 ms in the Miyake et al. study. Similarly, the raw score standard deviation of the antisaccade task was .15 in the present study and .07 in the Miyake et al. study. Also, the reliabilities for these two measures were higher in our study than theirs (respectively .90 vs. .72 for Stroop and .90 vs. .77 for antisaccade). It remains possible, however, that even though the standard deviations and reliabilities were higher in our study, one or both of these measures were mainly tapping something other than Inhibition in the older adult sample. A second possibility may rest in the fact that all of our Inhibition measures involved prepotent response inhibition, and none required resistance to proactive interference. Recall that the Hedden and Yoon (2006) study with older adults used indicators that tapped both prepotent response inhibition and resistance to proactive interference, and that study found that the two types of indicators loaded on different factors. Specifically, only the indicators targeting resistance to interference loaded on the Inhibition factor, whereas the inhibition of prepotent responses tasks loaded on a perceptual speed factor. Moreover, although the three Inhibition indicators in the Miyake et al. (2000) study (which, like the present study, only targeted inhibition of prepotent responses), did have significant paths to the Inhibition factor, the magnitude of significant relationships (Cohen, 1992) among the indicators in the Miyake et al. study was somewhat smaller for the Inhibition indicators ($M = .19$), as compared with the Shifting ($M = .29$) and Updating indicators ($M = .31$). In addition, verbal Stroop and antisaccade in the Miyake et al. study were significantly related to two of their Updating indicators, just as we found in our study. In this light, it seems possible that tasks targeting prepotent response inhibition may not be particularly robust measures of the Inhibition latent variable, particularly in older adults, and thus may have been inadequate to identify the factor in our relatively small sample of older adults.

A related question is whether the reduction of one factor for older relative to younger adults should be interpreted as support for the dedifferentiation hypothesis. Given the lack of overlap between the Updating and Shifting factors, we suggest that our results are not consistent with the dedifferentiation hypothesis. That is, dedifferentiation would predict that age-related cognitive decline is attributable to a single factor, which would in turn imply that correlations among indicator tasks and factors should increase dramatically with age. In contrast, our results show relatively little shared variance between the two factors, and the magnitude of the relationship between them is actually weaker than that revealed for young adults (Miyake et al., 2000) and is not significant. It is possible that this difference could at least partly arise from differences in specific measures and procedures. However, these differences were as minimal as possible, as we intentionally designed the present research to be analogous to the Miyake et al. study. Dedifferentiation would further suggest that all executive function indicators should have similar and very high loadings on a single factor. Again, this was not the case. The indicators for the Shifting factor were not significantly related to the Updating factor, nor were the indicators for Updating significantly related to the Shifting factor. Moreover, all but one of the simple indicator tasks for Shifting and Updating that were the same in this and the Miyake et al. study loaded on the same factors in each study, and the magnitude of path strengths from the indicators to their associated factors was similar in this study (.44 – .88) and the Miyake et al. study (.33 – .63).

The exception to the similarity in indicator loadings is the plus-minus task. Miyake et al. (2000) identified plus-minus as an indicator of Shifting ability in young adults, but plus-minus was robustly associated with the Updating factor in our study with older adults. Indeed, our results showed nonsignificant correlations between the plus-minus task and the two Shifting tasks (verbal and nonverbal local-global), but moderate to strong correlations with two of the Updating tasks (verbal and nonverbal keep-track). A likely reason for the correlation between plus-minus and the Updating indicators is that all these tasks require the active, concurrent maintenance of multiple task sets or representations. Specifically, these tasks require participants to keep the task set from the previous trial active in order to select the appropriate task set for the current trial (plus-minus), or maintain representations of previous trials in order to compare them to the current trial (keep-track and *N*-back tasks). In contrast, local-global tasks (i.e., Shifting indicators) require that participants use the cue provided on each trial to activate the appropriate task set for that trial and suppress information from previous trials (indeed, if the previous task set was different, i.e., a switch trial, its continued activation would contribute to interference, longer RTs, and more errors). Thus, Shifting requires one to be good at "forgetting" past responses/representations, whereas Updating and plus-minus tasks require one to be good at "remembering" them. It seems reasonable that these might rely on very different skills, and could also explain the negative trend between Updating and Shifting tasks in the present research. Moreover, because it seems likely that the concurrent activation feature of Updating tasks would place a higher demand on working memory capacity, and given that working memory capacity is thought to be generally reduced in older adults (Braver et al., 2001; Kray & Lindenberger, 2000; Mayr, 2001; Mayr &

Liebscher, 2001; West & Schwarb, 2006), it might make sense that a negative relationship between Shifting and Updating would be more evident among older than younger adults.

Factors Contributing to Complex Executive Function Task Performance

As the SEMs showed, the Updating factor is the most important predictor of performance for both TOH and WCST in this study. This is in contrast to the outcomes with young adults in the Miyake et al. (2000) study, which found that Shifting was the primary predictor for WCST, and Inhibition was primary for TOH. Although the complex tasks loaded differently across the two studies, it is worth noting that the magnitude of path estimates was similar. It is possible that the different outcomes could be a result of differences in some of the measures used to identify the underlying factors, but a more interesting possibility is that the reduction in working memory capacity in older adults (e.g., Braver et al., 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Just & Carpenter, 1992) causes an increased reliance on Updating skill, which seems to most directly represent working memory capacity (see Engle & Kane, 2004 and Friedman et al., 2006). For instance, the WCST requires participants to simultaneously maintain three different task sets (one each for color, number and shape), and efficient performance at the point of a task shift requires that one remember which task was previously appropriate and which has been ruled out on the basis of a particular test. The working memory requirements of the TOH are less evident; however, the task requires continuous maintenance of the rules for movement of objects and also memory for the series of moves that were previously unsuccessful in order to plan the success of future moves. Taken together with our outcomes from the plus-minus task and other evidence from the cognitive aging literature, we suggest the present results are consistent with the idea of an age-related difference in the reliance on Updating skills during complex executive function tasks (which place a particularly high demand on working memory capacity), perhaps as a direct consequence of an age-related decline in the ability to maintain the integrity of competing task sets in working memory and an overall reduction in working memory capacity (e.g., Braver et al., 2001). However, it remains for future longitudinal studies to establish this with certainty.

Verbal Versus Nonverbal Domains

Our results also show that verbal and nonverbal measures with the same basic structure (e.g., verbal and nonverbal keep-track) load on the same underlying factor—we found no evidence that nonverbal tasks are more related to each other than to verbal tasks within the same factor. This outcome may seem somewhat surprising given the apparent dissociation between the verbal and nonverbal domains (at least for tasks that tap the Inhibition factor) reported for patient ML (Hamilton & Martin, 2005). That is, there is a wealth of evidence indicating that when dissociations of functions are uncovered in a neurologically impaired individual, these functions represent separable components that play a role in the cognitive function of neurologically healthy individuals (for review see Rapp, 2001). In the case of ML, where there is a disruption of his verbal executive function ability but a sparing of his nonverbal executive function ability, it is likely that both processes also contribute separately in healthy, age-matched people. However, Hamilton and Martin have suggested that these dissociations may not be as apparent in healthy people because of a "shared dependence" of those tasks on neurotransmitters such as dopamine and norepinephrine (see also Stephan et al., 2003). That is, dependence on the same set of neurotransmitters may underlie the shared variance among these tasks, thus obscuring any effects of distinct verbal and nonverbal abilities or the neural substrates that support them. Thus, the apparent lack of separation in verbal and nonverbal ability in the present research cannot rule out the possibility that these are indeed separable factors.

Summary

The primary strength of this research is its relevance to the debate about the unity and diversity of executive functions in older adults. In sum, our results indicate that a single-factor solution is insufficient to explain executive function performance in older adults. As such, our outcomes are inconsistent with the dedifferentiation hypothesis, which predicts that executive function performance in older adults will be supported by a single factor, or at least on a few very strongly correlated factors. On the other hand, our results are consistent with the notion that the relative contributions of multiple underlying factors may undergo an age-related change. In particular, these outcomes may point toward important age-related differences in the role of Updating abilities, perhaps as a function of age-related reductions in working memory capacity that particularly manifest in reduced efficiency at maintaining the integrity of competing mental representations. We suggest that this decline leads to an increased reliance on Updating skill for executive function tasks, especially those that place a particularly high demand on working memory capacity. A remaining question concerns the relationship between Updating skills and intelligence, and whether and how it may change with age. It is hoped that the present study will offer new directions for future research to expand and integrate cognitive theories of executive function across the life span.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131,* 567–589.

Antsey, K., Hofer, S., & Luszcz, M. (2003). Cross-sectional and longitudinal patterns of dedifferentiation in late-life cognitive and sensory function: The effects of age, ability, attrition, and occasion of measurement. *Journal of Experimental Psychology: General, 132,* 470–487.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology, 8,* 234–238.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs, 23,* 191–234.

Balota, D., & Faust, M. (2001). Attention in dementia of the Alzheimer's type. In F. Boller and F. S. Cappa (Eds.), *Handbook of Neuropsychology, 2nd Ed.* (pp. 51–79). Amsterdam: Elsevier.

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

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., et al. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130,* 746–763.

Brennan, M., Welsh, M. C., & Fisher, C. B. (1997). Aging and executive function skills: An examination of a community-dwelling older adult population. *Perceptual & Motor Skills, 84,* 1187–11987.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Chiappe, R., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28,* 8–17.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good enough" interpretations of garden path sentences. *Discourse Processes, 42,* 205–238.

Clark, H. H., & Brownwell, H. H. (1975). Judging up and down. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 339–352.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning & Motivation, 44,* 145–199.

Fernandes, M. A., Pacurar, A., Moscovitch, M., & Grady, C. (2006). Neural correlates of auditory recognition under full and divided attention in younger and older adults. *Neuropsychologia, 44,* 2452–2464.

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

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172–179.

Gevins, A. S., & Cutillo, B. C. (1993). Neuroelectric evidence for distributed processing in human working memory. *Electroencephalography in Clinical Neurophysiology, 87,* 128–143.

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging, 18,* 696–713.

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 1–13.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher and A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual* – Revised and Expanded. Lutz, FL: Psychological Assessment Resource.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology, 20,* 511–528.

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability and validity of two executive function tests. *Assessment, 4,* 249–257.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, Whole No. 89.*

Johnson, W., & Bouchard, T. (2005). The structure of human intelligence: It's verbal, perceptual, and image rotation. *Intelligence, 33,* 393–416.

Jöreskog, K., & Sörbom, D. (2001). *LISREL 8: User's reference guide.* Lincolnwood, IL: Scientific Software International.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kane, M., & Engle, R. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9,* 637–671.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *NeuroReport, 6,* 3581–3585.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15,* 126–147.

Lu, C. H., & Proctor, R. W. (1994). Processing of an irrelevant location dimension as a function of the relevant stimulus dimension. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 286–298.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain & Language, 89,* 76–82.

Martin, R. C., & Lesch, M. R. (1996). Associations and dissociations between language impairment and list recall: Implications for models of short-term memory. In S. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 149–178). Hove, U. K.: Erlbaum.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition, 27,* 759–767.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging, 16,* 96–109.

Mayr, U., & Leibscher, T. (2001). Is there an age deficit in the selection of mental sets? *European Journal of Cognitive Psychology, 13,* 47–69.

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

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6,* 328–362.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology, 9,* 491–500.

Rapp, B. (2001). *The Handbook of Cognitive Neuropsychology.* Philadelphia: Psychology Press.

Roberts, R. J., Jr., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

Rvan, J. J., & Ward, L. C. (1999). Validity, reliability, and standard errors of measurement for two seven-subtest short forms of the Wechsler Adult Intelligence Scale-III. *Psychological Assessment, 11,* 207–211.

Salthouse, T. (2001). Structural models of the relations between age and measures of cognitive function. *Intelligence, 29,* 93–115.

Salthouse, T., Atkinson, T., & Berish, D. (2003). Executive function as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132,* 566–594.

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

Smith, E., Jonides., J., Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6,* 11–20.

Stephan, K. E., Marshall, J. C., Friston, K. J., Rowe, J. B., Ritzi, A., Zilles, K., et al. (2003). Lateralized cognitive processes and lateralized task control in the human brain. *Science, 301,* 384–386.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Wechler, D. (1997). *Weschsler Adult Intelligence Scale-III.* San Antonio, TX: The Psychological Corporation.

West, R. (2004). The effects of aging on controlled attention and conflict processing in the Stroop task. *Journal of Cognitive Neuroscience, 16,* 103–113.

West, R., & Bowry, R. (2005). Aging, working memory, and prospective memory. *Psychophysiology, 42,* 698–712.

West, R., & Schwarb, H. (2006). The influence of aging and frontal function on the neural correlates of regulative and evaluative aspects of cognitive control. *Neuropsychology, 20,* 468–481.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors, 5,* 7–17.

Zook, N., Welsh, M., & Ewing, V. (2006). Performance of healthy, older adults on the Tower of London-Revised: Associations with verbal and nonverbal abilities. *Aging, Neuropyschology, and Cognition, 13,* 1–19.

Received: April 10, 2007
Revision received: February 6, 2008
Accepted: February 8, 2008 ■