Journal of Experimental Psychology: General
2003, Vol. 132, No. 4, 566–594

0096-3445/03/$12.00 DOI: 10.1037/0096-3445.132.4.566

Executive Functioning as a Potential Mediator of Age-Related Cognitive Decline in Normal Adults

Timothy A. Salthouse, Thomas M. Atkinson, and Diane E. Berish
University of Virginia

Critical requirements for the hypothesis that executive functioning is a potential mediator of age-related effects on cognitive functioning are that variables assumed to reflect executive functioning represent a distinct construct and that age-related effects on other types of cognitive functioning are reduced when measures of executive functioning are statistically controlled. These issues were investigated in a study involving 261 adults between 18 and 84 years of age. Although age-related effects on various cognitive abilities were substantially reduced after statistical control of the variance in measures hypothesized to represent executive functioning, there was only weak evidence for the existence of distinct constructs corresponding to executive functioning or to aspects of executive control concerned with inhibition, updating, or time sharing.

In recent years there has been considerable interest within neuropsychology and in certain areas of cognitive psychology in the concept of executive functioning, broadly defined as control processes responsible for planning, assembling, coordinating, sequencing, and monitoring other cognitive operations. The executive functioning concept is loosely based on an analogy to a business executive who is not necessarily a specialist in any particular domain but instead is responsible for supervising and managing many different domains. Because disruption of central control processes might result in the impairment of relevant behavior even if the component processes are intact, executive functioning clearly has the potential to affect performance in a wide variety of cognitive variables.

The concept of executive functioning could be informative about adult age differences in cognitive functioning for at least two reasons. First, executive functioning encompasses concepts such as inhibition, working memory (updating), and attentional capacity (time sharing) that have played a prominent role in theories of cognitive aging. However, past research has largely viewed these concepts as independent of one another and has not examined the possibility that they represent different aspects of a single construct. And second, executive functioning is often associated with the frontal lobes, and in fact, the terms *frontal functions* and *executive functions* are frequently used interchangeably in the literature. Executive functioning is therefore relevant to the *frontal hypothesis of aging*, in which many age-related cognitive deficits are postulated to be associated with a deterioration of the frontal lobes (e.g., Albert & Kaplan, 1980; Andres & van der Linden, 2001; Crawford, Bryan, Luszcz, Obonsawin, & Stewart, 2000; Daigneault, Braun, & Whitaker, 1992; Dempster, 1992; Mittenberg, Seidenberg, O'Leary, & DiGiulio, 1989; Moscovitch & Winocur, 1995; Parkin, 1997; Perfect, 1997; Phillips & Della Sala, 1998; Raz, 2000; Raz, Gunning-Dixon, Head, Dupuis, & Acker, 1998; Schretlen et al., 2000; Shimamura, 1984; Souchay, Isingrini, & Espagnet, 2000; Troyer, Graves, & Cullum, 1994; Veroff, 1980; West, 1996; Whelihan & Lesher, 1985). It is probably unrealistic to expect a precise one-to-one correspondence between executive functioning and discrete neuroanatomical structures (e.g., Anderson, Damasio, Jones, & Tranel, 1991; Foster, Black, Buck, & Bronskill, 1997; Mountain & Snow, 1993; Reitan & Wolfson, 1994, 1995; Robbins et al., 1997; Rubin, 1999; Shallice & Burgess, 1991; Tranel, Anderson, & Benton, 1994), but the frontal lobes, and particularly the prefrontal cortex, have been implicated as an important substrate for executive processes in studies of patients with focal lesions and in neuroimaging research.

It is not surprising in light of the hypothesized importance of executive functioning as a theoretical construct, and its presumed relevance to age differences in cognitive functioning, that researchers have speculated that it may be responsible for mediating at least some of the age-related effects on different measures of cognitive functioning (e.g., Bryan, Luszcz, & Pointer, 1999; Crawford et al., 2000; Ferrer-Caja, Crawford, & Bryan, 2002; Troyer et al., 1994). However, two conditions must be satisfied before executive functioning can be considered plausible as a mediator of age-related influences in cognitive functioning. The first is that it is meaningful to refer to executive functioning as a distinct construct, and the second is that statistical control of the variance in measures of executive functioning is associated with a decrease in the age-related effects in the relevant measures of cognitive functioning. The rationale for each of these conditions is elaborated in the following paragraphs.

The Executive Functioning Construct

Despite frequent references to the concept of executive functioning within the neuropsychological literature, executive pro-

Timothy A. Salthouse, Thomas M. Atkinson, and Diane E. Berish, Department of Psychology, University of Virginia.

This research was supported by National Institute on Aging Grant RO1AG19627 to Timothy A. Salthouse. We thank Emilio Ferrer-Caja, D. Z. Hambrick, Chris Hertzog, and Jack McArdle for valuable suggestions and Jason Chin, Sophy Cogill, Meredith Haskins, Chris Makarsky, Jane Mendle, Lamont Mitchell, and Monica Smith for scheduling and testing participants.

Correspondence concerning this article should be addressed to Timothy A. Salthouse, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, Virginia 22904-4400. E-mail: salthouse@virginia.edu

cesses are very complex and have proven difficult to assess and understand. To illustrate, Baddeley (1998) suggested that they “are probably the most complex aspects of human cognition” (p. 525), and Levine, Stuss, and Milberg (1995) claimed that “the assessment of executive functioning may be the most challenging task faced by neuropsychologists” (p. 755). This complexity is apparent in the following characterizations of executive functioning from recent handbooks in neuropsychology:

> The executive functions consist of those capacities that enable a person to engage successfully in independent, purposive, self-serving behavior. (Lezak, 1995, p. 42)

> Executive functions include the following abilities: 1. Formulating goals with regard for long-term consequences. 2. Generating multiple response alternatives. 3. Choosing and initiating goal-directed behaviors. 4. Self-monitoring the adequacy and correctness of the behavior. 5. Correcting and modifying behaviors when conditions change. 6. Persisting in the face of distraction. (Malloy et al., 1998, p. 574)

> The term executive function . . . [is a] shorthand description of a multidimensional construct referring to a variety of loosely related higher-order cognitive processes including initiation, planning, hypothesis generation, cognitive flexibility, decision making, regulation, judgment, feedback utilization, and self-perception that are necessary for effective and contextually appropriate behavior. (Spreen & Strauss, 1998, p. 171)

These are obviously quite broad conceptualizations, and one might wonder how such diverse capacities, abilities, or processes might be assessed. Among the tests most frequently used by neuropsychologists to evaluate executive functioning, and that were included in the current study, are the following.

Wisconsin Card Sorting Test (WCST)

This test is presumed to require maintenance of task set, flexibility in response to feedback, and avoidance of perseverative tendencies, perhaps by inhibiting the prior response category when it is no longer appropriate. The WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993) was recently described by Delis, Kaplan, and Kramer (2001) as “the gold standard of executive function tests” (p. 2). Adult age differences on one or more measures of performance in this test have been reported numerous times (e.g., T. Y. Arbuckle & Gold, 1993; Axelrod & Henry, 1992; Boone, Miller, & Lesser, 1993; Bryan et al., 1999; Crawford et al., 2000; Daigneault et al., 1992; Davis et al., 1990; Fristoe, Salthouse, & Woodard, 1997; Haaland, Vranes, Goodwin, & Garry, 1987; Head et al., 2002; Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Liben et al., 1994; Mejia et al., 1998; Parkin & Java, 1999, 2000; Parkin & Lawrence, 1994; Parkin & Walter, 1991, 1992; Parkin, Walter, & Hunkin, 1995; Salthouse, Fristoe, & Rhee, 1996; Salthouse, Toth, Hancock, & Woodard, 1997; Souchay et al., 2000).

Verbal and Figural Fluency

Fluency tests are presumed to require maintenance of a task set (i.e., report only items satisfying the designated constraints), generation of multiple response alternatives, monitoring of responses to avoid repetitions, and use of different retrieval strategies. Although the magnitude of age differences in verbal fluency tests may be moderated by level of vocabulary (e.g., Salthouse, 1993; Salthouse, Hancock, Meinz, & Hambrick, 1996), there are many reports of significant age differences in both verbal and figural fluency tests (e.g., T. Y. Arbuckle & Gold, 1993; Ardila & Rosselli, 1989; Baltes & Lindenberger, 1997; Birren, 1955; Bolla, Lindgren, Bonaccorsy, & Bleecker, 1990; Bryan et al., 1999; Isingrini & Vazou, 1997; Keys & White, 2000; Maki, Zonderman, & Weingartner, 1999; McCrae, Arenberg, & Costa, 1987; Mejia, Pineda, Alvarez, & Ardila, 1998; Mittenberg et al., 1989; Parkin & Java, 2000; Parkin & Lawrence, 1994; Parkin & Walter, 1991, 1992; Parkin et al., 1995; Phillips, 1999; Ronnlund, Lovden, & Nilsson, 2001; Salthouse, Fristoe, & Rhee, 1996; Salthouse et al., 1997; Souchay et al., 2000; Tomer & Levin, 1993; Whelihan & Lesher, 1985).

Trail Making

Various types of trail-making tests are presumed to require cognitive flexibility to efficiently switch between sequences when appropriate. In order to isolate aspects of switching and maintenance of position in each sequence from more general processes concerned with perceiving and responding, the difference in performance between Version B (with alternating number and letter targets) and Version A (with only number targets) is frequently used as the critical variable in this task. Age differences favoring younger adults have consistently been reported in either the B-A difference or the time to complete Version B of the Trail Making Test (e.g., T. Y. Arbuckle & Gold, 1993; Keys & White, 2000; May & Hasher, 1998; Rasmussen, Zonderman, Kawas, & Resnick, 1998; Salthouse, Fristoe, & Rhee, 1996; Salthouse et al., 1997) or in a variant of it referred to as the Connections Test (Salthouse et al., 2000).

Tower of Hanoi and Tower of London

These tests are presumed to require the ability to develop and execute a plan, consisting of a sequence of steps, to convert an initial configuration of disks on towers into a final configuration. Increased age has been found to be associated with poorer performance in several different types of tower tests (e.g., Allamanno, Della Sala, Laiacona, Pasetti, & Spinnler, 1987; Andres & van der Linden, 2001; Bisiacchi, Sgaramella, & Farinello, 1998; Brennan, Welsh, & Fisher, 1997; Charness, 1987; Crawford et al., 2000; Gilhooly, Phillips, Wynn, Logie, & Della Salla, 1999; Head et al., 2002; Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999; Robbins et al., 1998; Ronnlund et al., 2001; Vakil & Agmon-Ashinazi, 1997).

Although one might conclude on the basis of the existing results that there are age differences in executive functioning, such a conclusion may actually be premature. What is needed before this conclusion would be justified is evidence that the various tasks reflect the same theoretical construct, and that the construct is distinct from other constructs.

The issue of what tasks used to assess executive functioning actually represent was discussed by Duncan, Johnson, Swales, and Freer (1997) in the following passage:

> One may ask, however, how much such very different tasks have in common. In truth, the only known common factor is that these tests have sometimes shown significant differences between groups of frontal patients and normal controls; a weak criterion given the very broad disorganization of behaviour that can sometimes follow frontal lesions. (p. 714)

Most tasks hypothesized to assess executive functioning can be claimed to have moderate levels of face validity, but face validity provides a weak foundation for evaluating what the variables represent because it is ultimately based on subjective judgments. Moreover, the absence of an external criterion of executive functioning against which the variables can be validated has made it impossible to assess predictive validity.

One aspect of validity that can be investigated is whether different variables represent the same construct, and several methods can be used for this purpose. For example, the variables could be compared with respect to their susceptibility to experimental manipulations, or to focal brain damage. In each case, two or more variables can be inferred to assess the same construct to the extent that they have similar patterns of sensitivity and specificity. Sensitivity would be demonstrated if the target variables are affected by the same manipulation, or by brain damage to the same region, and specificity would be established if the variables are not affected by other manipulations, or by damage to different brain regions. A similar rationale can be applied in individual-differences research by viewing individuals as analogous to different experimental treatments or manipulations. The analogues to sensitivity and specificity in this context are convergent and discriminant validity, because the former refers to the extent to which variables presumed to measure the same construct covary together across people, and the latter corresponds to a lack of covariation between variables presumed to reflect different constructs. Because it is not always feasible to use the same experimental manipulations with different types of tasks, and because ethical considerations preclude control over the size and location of brain lesions in humans, the individual-differences approach is often the most practical method of investigating construct validity. The key assumption in this method is that if a set of variables reflects the same dimension of human variation, then these variables should have moderate to large correlations with each other (i.e., convergent validity) but much weaker correlations with variables representing other constructs (i.e., discriminant validity).

Unfortunately, evidence of construct validity for the executive functioning construct has often been weak, because correlations among variables from "executive" tasks have frequently been low relative to their correlations with variables from "nonexecutive" tasks (e.g., Andres & van der Linden, 2001; Boone, Panton, Gorsuch, Gonzalez, & Miller, 1998; Daigneault et al., 1992; Della Sala, Gray, Spinnler, & Trivelli, 1998; Duncan et al., 1997; Kafer & Hunter, 1997; Kopelman, 1991; Kramer et al., 1994; Lehto, 1996; Obonsawin et al., 2002; Rabbitt, 1997; Robbins et al., 1997, 1998; Salthouse, Hancock, et al., 1996; but see Hanes, Andrewes, Smith, & Pantelis, 1996, for an exception). Furthermore, the failure to find a systematic pattern of higher within-construct than between-constructs correlations is not simply attributable to the use of relatively crude summary scores from neuropsychological tests, because Duncan et al. (1997) examined more specific measures of performance in each task and still failed to find a pattern of correlations consistent with the existence of a distinct executive functioning construct.

Results from a recent test battery explicitly designed to assess executive functioning (the Delis–Kaplan Executive Function System; Delis et al., 2001) can be used to illustrate this prototypical pattern. The data used to establish the norms for this test battery are particularly valuable because the sample was moderately large (i.e., 875 adults), and attempts were made to ensure that it was representative of the U.S. population. Furthermore, many tests mentioned in discussions of executive functioning were included in the battery.

Correlations were reported among summary scores for sorting, tower, letter fluency, and figural fluency tasks for 425 adults between 20 and 49 years of age and for 450 adults between 50 and 89 years of age. Variants of the trail-making and Stroop tasks were also included in the test battery, but correlations were not reported for measures of differences between alternating and same sequences or for differences between incongruent and either congruent or neutral conditions, and because the variables that were reported may represent a mixture of executive and nonexecutive processes, they were not included in the following analyses. Median correlations among the four executive function variables mentioned above were .25 for 20- to 49-year-olds and .33 for 50- to 89-year-olds. These relatively low values suggest that the executive functioning construct in this test battery had weak convergent validity. Correlations were also reported for a sample of 292 individuals with another type of test, the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 2000) of episodic memory, and they were similar in magnitude to those among the purported executive measures. To illustrate, correlations with a measure of recall across Trials 1 to 5 were .28 for sorting, .23 for tower, .32 for letter fluency, and .27 for design fluency. Because these values are within the same range as the correlations of the executive function variables with one another, these normative data provide weak evidence for discriminant validity of the executive functioning construct.

At least three interpretations can be proposed to account for the low correlations found among variables hypothesized to reflect executive functioning. First, some of the variables may have low reliability, and relations with other variables are necessarily limited if only a small proportion of the variance in the variables is systematic and available to be associated with other variables. In fact, some researchers have suggested that executive functioning measures should not be expected to be reliable because they are designed to assess the ability to cope with novel problems, and the problems are no longer novel after the first administration (e.g., Burgess, 1997; Denckla, 1995; Lowe & Rabbitt, 1997; Phillips, 1997; Rabbitt, 1997).[1] Another factor related to low reliability that could be contributing to the weak correlations is the use of observed variables that contain measurement error and that likely represent other factors in addition to the construct of interest. Stronger estimates of the relations between the constructs of interest might therefore be obtained by examining relations between latent constructs that represent the reliable variance that several variables have in common.

A second possible interpretation of the low correlations is that the influence of executive functioning on most variables could be small relative to other types of influences. That is, because executive functioning typically refers not to a single discrete cognitive operation but rather to the coordination of multiple processes or operations, it may be difficult to distinguish the influence of

[1] An intriguing implication of the interpretation that executive functioning variables change their meaning after the first assessment is that the pattern of correlations with other variables should differ from the first to the second administration of the tests. Unfortunately, there is apparently no evidence relevant to this speculation at the current time.

executive functioning from the efficiency of constituent processes. The role of executive functioning may also be rather limited in many laboratory tasks because much of the organization or structure of the tasks is provided by the experimenter and does not need to be discovered or created by the research participant (e.g., Burgess, 1997; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Lezak, 1995; Malloy, Cohen, & Jenkins, 1998; Shallice & Burgess, 1991; Spreen & Strauss, 1998; Tranel et al., 1994).

Finally, correlations among variables hypothesized to reflect executive functioning may be low either because the variables might not represent a separate construct in normal adults or because they reflect different types of executive functions rather than a unitary construct. The first interpretation is essentially that there may not be a distinct dimension corresponding to executive functioning in the variation apparent among normal healthy adults, whereas the second is that there may be several different types of executive functioning. Delis et al. (2001) appear to have accepted this second interpretation because they suggested that "the relatively low positive correlations between . . . [executive function] tests indicate that the instruments are not interchangeable and measure unique aspects of executive functioning with some overlap of variance" (p. 82). Various fractionations of executive functions have been discussed (e.g., Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Della Sala et al., 1998; Eslinger & Grattan, 1993; Lehto, 1996; Miyake et al., 2000; Parkin, 1997; Parkin & Lawrence, 1994), but there is not yet a consensus regarding the number or nature of separable components.

One means of investigating the possibility of multiple executive functions is to focus on theoretical processes hypothesized to be involved in executive functioning instead of relying on global scores derived from complex neuropsychological tests. Indeed, cognitive psychologists have tended to emphasize control processes assumed to be responsible for assembling, coordinating, scheduling, and monitoring other cognitive processes and on whatever is responsible for focusing, shifting, dividing, and inhibiting attention. This perspective was articulated by Miyake et al. (2000) in their characterization of executive functions as "general-purpose control mechanisms that modulate the operation of various cognitive subprocesses, and thereby regulate the dynamics of human cognition" (p. 50).

Because the number of possible control processes is very large, many different types of tasks and variables could be (and have been) postulated to assess aspects of executive control from the cognitive process perspective. However, concerns about validity can also be raised with respect to variables assumed to reflect executive control processes because most researchers have relied on face validity judgments that a particular variable represents an aspect of executive control but have seldom provided evidence of any other type of validity. This is unfortunate, because it is desirable that validity be empirically established for all variables, even those derived from elegant cognitive tasks designed to provide relatively direct assessments of a single cognitive process. That is, regardless of how "pure" a variable may appear, if it truly reflects the theoretical construct of interest then it should be expected to be at least moderately related to other variables hypothesized to represent the same construct and much less strongly related to variables hypothesized to represent other constructs.

A delicate balance often exists between the narrowness and breadth of theoretical constructs. For example, very narrow constructs, perhaps defined by variables from the same paradigm but involving slightly different stimulus material, could be dominated by method- and measure-specific influences rather than by influences associated with the theoretical construct of interest. In contrast, if the construct is defined on the basis of a diverse collection of variables involving different methods, materials, and measures, it could become so broad that its meaning is obscure. There is no ideal solution to this dilemma, and perhaps the best that can be done is to examine empirical patterns with different combinations of variables to determine the level of aggregation that is most meaningful.

Three constructs hypothesized to be aspects of executive control were investigated in the current study: inhibition of prepotent responses, updating of continuously changing information, and time sharing between two simultaneous activities. Other aspects could clearly have been investigated, but there are practical limitations on the number that can be examined in a single study. Furthermore, these particular constructs may be more tractable than broader constructs such as working memory, for which there currently appears to be little consensus. To illustrate, Kimberg, D'Esposito, and Farah (1997) claimed that "if you ask 100 cognitive psychologists what working memory is, you will get 100 different answers" (p. 188), and their suggestion seems to have been confirmed by the diversity of views expressed by the 10 contributors to a volume on working memory edited by Miyake and Shah (2000).

Efficiency of inhibiting or suppressing prepotent responses can be hypothesized to represent one aspect of executive functioning, because it is frequently necessary to overcome reliance on habit, familiarity, or processes that are already underway. Some responses are dominant because they represent a highly overlearned habit such as reading, others because they correspond to a reflexive shift of attention to external stimuli, and others because the response is already in preparation. However, in each case it is sometimes necessary to inhibit or suppress these responses, and the efficiency with which this is accomplished can be presumed to reflect effectiveness of executive functioning. Four variables were used to assess aspects of inhibiting prepotent responses in the current study.

First, the relation between color naming time in Stroop incongruent (i.e., color words written in colors different from the words) and neutral (i.e., a string of colored *X*s) conditions can be hypothesized to reflect the efficiency of suppressing a prepotent word-reading response, distinct from other processes involved in perceiving and naming items, with a larger difference corresponding to weaker inhibition. Second, the relation between reading time and comprehension with distracting material interpolated in the passage and reading time and comprehension with normal material can be hypothesized to reflect the efficiency of suppressing the tendency to read irrelevant words, distinct from other processes involved in perceiving and reading text, with a larger difference corresponding to weaker inhibition. Third, the relation between time or accuracy of responding to a target stimulus when an irrelevant flash occurs in the same (congruent) or opposite (incongruent) side of the computer screen can be hypothesized to be attributable to the reflexive tendency to shift attention in the direction of the irrelevant flash, with a larger difference in reaction time and/or accuracy corresponding to less effective inhibition or suppression of the reflexive tendency. And fourth, the probability of stopping a response that is in the process of being executed can be hypothesized to reflect the effectiveness of suppressing or inhibiting cognitive operations, with a higher stopping probability corresponding to more effective inhibition.

At least two earlier studies have reported correlations between the Stroop interference measure and the reading-with-distraction difference score, but both were quite low. Stoltzfuz, Hasher, Zacks, Ulivi, and Goldstein (1993) found a correlation of .12 in a sample of college students, and Earles et al. (1997) found a correlation of .20 in a sample of 266 adults between 20 and 90 years of age. One possible reason for these low correlations is high specificity of the variables, because low correlations among the interference scores derived from different versions of the Stroop task were reported by Schilling, Chetwynd, and Rabbitt (2002) and Ward, Roberts, and Phillips (2001). However, this need not always be the case, as moderate correlations (ranging from .28 to .52, and from .42 to .72 after adjusting for reliability) were reported in a study by Salthouse and Meinz (1995).

A second hypothesized aspect of executive functioning involves updating internal representations of external information because environmental information is seldom static and effective cognitive functioning requires accurate monitoring of the status of continuously changing information. Four tasks were used to assess efficiency of updating in this study. In the keeping track task, a series of words representing exemplars from four different categories was displayed on the screen, followed by a probe containing a category name and an exemplar. The task for the participant was to decide whether the exemplar was the most recently presented item from that category. Successful performance can be assumed to require continuous replacement of category–exemplar associations. In the digit-monitoring task, a series of digits was presented with instructions to make one response to nontargets and a different response to targets, which were defined as every third odd digit. Successful performance can be hypothesized to require monitoring and revising the count of odd digits while also registering and responding to each digit. The matrix-monitoring task involved a series of displays beginning with a dot in one cell of a matrix, followed by a sequence of two arrows, and finally by a dot in another cell of the matrix. The task was to decide whether the dot was in the cell that would have resulted from movement in the directions indicated by the arrows. Successful performance can be assumed to require accurate revision of the dot position with each successive arrow. Finally, the *n*-back task consisted of a series of successively presented digits, with the research participant instructed to repeat the items that occurred *n* items back in the sequence (where *n* ranged from 0 to 2). Successful performance can be presumed to require constant exchange of the new item, the response item, and any stored items.

Although a number of studies have reported moderate correlations among performance measures from various types of updating tasks (e.g., Kyllonen & Christal, 1990; Miyake et al., 2000; Salthouse, Babcock, & Shaw, 1991; Salthouse, Hancock, et al., 1996), variables from the particular tasks described above have apparently not been examined in the same study.

A third hypothesized aspect of executive functioning involves coordination of two or more concurrent activities. That is, many complex cognitive tasks require that several cognitive operations be carried out nearly simultaneously, and thus an important dimension of executive functioning concerns the efficiency with which these operations can be coordinated and scheduled. Effectiveness of time sharing can be evaluated by comparing performance when two tasks are performed together and when they are performed alone. Three pairs of tasks were used to assess the efficiency of time sharing in this study. One combination consisted of a visual–motor tracking task and an auditory–verbal paired-associates task. The primary task involved listening to six pairs of words and then recalling the second member of the pair when presented with the first. The secondary task consisted of continuous tracking of a randomly moving target on a computer screen by manipulation of a track ball. A second pair of tasks consisted of a driving task and the *n*-back updating task. The primary task was to listen to a sequence of 25 digits and repeat the items after a lag of 0, 1, or 2 items. The secondary task was a simulated driving task in which a steering wheel was used to keep a "vehicle" in the middle of a curved "road." The third pair of tasks consisted of connecting letters in alphabetical sequence and counting backward. The primary task was to count backward by threes as quickly and accurately as possible from a number read by the examiner. The secondary task was to draw lines to connect circled letters in alphabetical sequence as rapidly as possible.

This particular combination of dual tasks has not previously been examined in a single study. However, Salthouse and Miles (2002) recently found moderate correlations (i.e., .36 to .68) among dual-task difference scores when the visual–motor tracking task was performed concurrently with a paired-associates task, a series completion task, and a spatial directions task.

Investigation of Construct Validity

The current study relied on the individual-differences approach to the investigation of construct validity with a combination of neuropsychological, cognitive process, and psychometric variables. A schematic representation of the neuropsychological and cognitive process approaches to the assessment of executive functioning is portrayed in Figure 1. Above the dashed line are variables used to assess executive functioning from the neuropsychological tradition, and below are variables used to assess aspects of executive control from the cognitive process tradition. Figure 1 can be interpreted as representing the hypothesis that variables frequently categorized as representing the same construct are systematically related to one another. That is, the circle represents the theoretical construct that is presumed to influence the variables to which it is linked with arrows, and the relation between any pair of variables corresponds to the square root of the product of the coefficients for the paths from the construct to those variables. However, the presence of correlations among the target variables is relevant only to the convergent validity aspect of construct validity, and in order to establish discriminant validity it must be demonstrated that the target variables are much less strongly related to variables representing other constructs.

Figure 2 portrays four models of how construct validity might be assessed. The illustration in Panel A is similar to that in Figure 1 in that only convergent validity can be evaluated by determining the relations among the candidate variables. One means of investigating discriminant validity involves examining the relations between the hypothesized construct and constructs representing other abilities. This is portrayed in Panel B in the form of the dotted lines connecting the target construct (on the left) with other constructs (on the right). For the target construct to be considered distinct from other constructs, its correlations to the other constructs should be substantially less than 1.0. The model portrayed in Panel C differs from that in Panel B by allowing the candidate

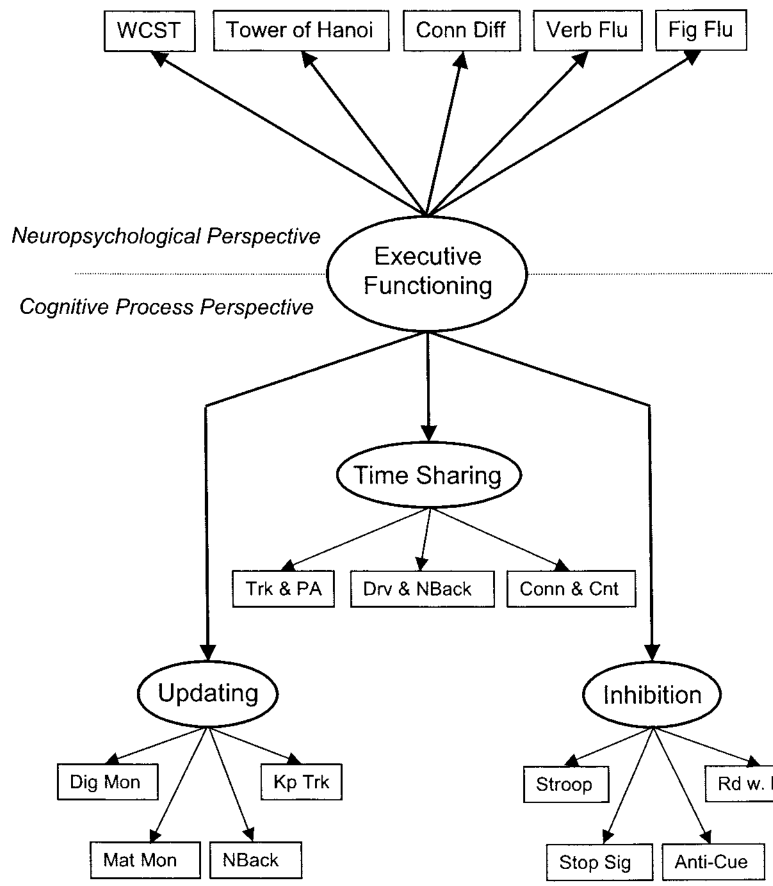

Figure 1. Relation of observed variables to hypothesized executive functioning constructs. WCST = Wisconsin Card Sorting Test; Conn Diff = connections (difference); Verb Flu = verbal fluency; Fig Flu = figural fluency; Trk & PA = tracking with paired associates; Drv & NBack = driving with *n*-back; Conn & Cnt = connect and count; Dig Mon = digit monitoring; Kp Trk = keeping track; Mat Mon = matrix monitoring; Rd w. Dist = reading with distraction; Stop Sig = stop signal.

variables to have loadings on other constructs if they contribute to improved fit of the model to the data. As in the model in Panel B, discriminant validity can be examined by inspecting the correlations with other constructs, and convergent validity can be evaluated by examining the extent to which the target variables are related to the hypothesized construct after allowing relations between them and other constructs.

The most demanding test of construct validity is represented in Panel D, in which the relations of the candidate variables to other constructs are taken into consideration when evaluating the degree to which the variables have variance in common. This model provides a strong test of construct validity because the influence of other constructs is examined at the level of individual variables and not simply at the level of the construct, as in the model portrayed in Panel B. Convergent validity in this model would be established by a discovery of significant amounts of residual variance shared among the candidate variables, and discriminant validity would be established by a finding of weak relations of the individual variables to constructs other than the hypothesized construct.

Model D is also interesting because the results from this type of model can be expected to be relevant to claims that a particular variable reflects one or more processes that are important determinants of a broader ability. For example, it has been postulated that certain variables, such as those from goal neglect (Duncan et al., 1996), antisaccade (Kane, Bleckley, Conway, & Engle, 2001), and context maintenance (Braver et al., 2001) tasks, represent processes that are fundamental to fluid intelligence. However, these claims might be questioned if the critical variable were found to be related to several different abilities (because the association of the variable with the target ability would not be unique) or if other variables were found to have relations of similar magnitude to the ability (because the association to the ability would not be exclusive to the hypothesized critical variable).

Each of the models of construct validity portrayed in Figure 2 was investigated in the current study with neuropsychological variables used to assess executive functioning and with variables hypothesized to assess processes of inhibition, updating, and time sharing. A long history of psychometric research has established the existence of different cognitive abilities (e.g., Carroll, 1993), and consequently a battery of cognitive tests was administered to assess cognitive abilities of fluid intelligence (gF), episodic memory, perceptual speed, and vocabu-

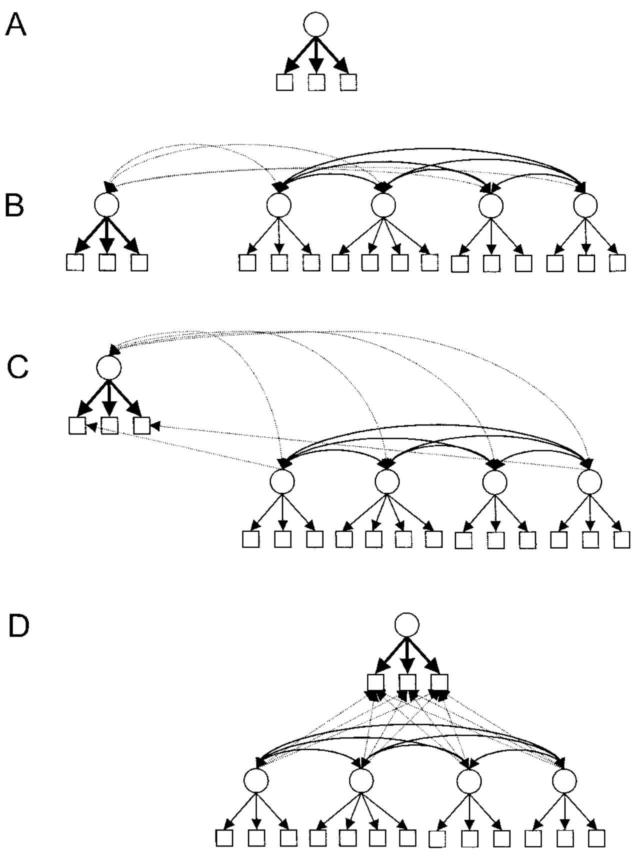

Figure 2. Four alternative models for the assessment of construct validity. The boxes represent measured variables, and the circles represent latent constructs. Convergent validity is supported when the relations represented by the thick lines are large, and discriminant validity is supported when the relations represented by the dotted lines are small.

lary, to serve as the other constructs when evaluating the models in Panels B, C, and D.

A recent study by Miyake et al. (2000) has some similarities to the current study in that it combined neuropsychological and cognitive process approaches to the assessment of executive functioning. However, their sample consisted of college students with a limited age range, and likely also with a relatively narrow range of cognitive ability. Furthermore, although estimates of reliability were reported for the cognitive process variables, some were rather low (e.g., .31 and .42), and many of the variables had weak loadings on the hypothesized constructs (i.e., .33 to .63, with a median of .46). Because this indicates that the constructs were associated with only about 20% of the variance in each variable, the majority of the variance in the variables was unaccounted for in their model. The authors concluded that three latent constructs postulated to represent executive functioning aspects corresponding to switching, inhibition, and updating exhibited discriminant validity because the correlations among them were only moderate (i.e., .42 to .63). However, they did not examine whether the executive functioning constructs were distinct from already established cognitive ability constructs, nor did they investigate the model represented in Panel D of Figure 2.

Mediation of Age–Cognition Relations

Figure 3 portrays a simple mediation model in which at least some of the age-related influences on the cognitive construct are indirect and operate through the mediator. (Because some variables may have nonlinear age relations, the model includes an age-squared variable to represent quadratic relations between age and the constructs.) For mediation to be plausible, not only should the variables hypothesized to correspond to the mediator covary together in a manner consistent with a single dimension of human variation, but the direct relation of age on the cognitive construct should be small relative to the total age relation. Stated somewhat differently, if mediation is operating then statistical control of the mediator should result in a decrease in the age-related effects on other cognitive variables. The rationale is that if at least some of the effects of aging on variable Y (e.g., physical strength) are mediated through effects on variable X (e.g., muscle mass), then effectively equating everybody on variable X should reduce the relation between age and variable Y. A finding of substantial reduction in the age-related variance is not sufficient to establish that executive functioning is a causal mediator of those effects, because there are a number of alternative interpretations that could account for this type of reduction in variance (e.g., the causal direction could be reversed, or a third variable could be responsible for effects on both variables). However, a pattern of attenuated age–cognition relations is necessary for executive functioning to be plausible as a mediator, because if the age-related variance in other variables was not reduced after control of the variance in the executive functioning construct, then the two sets of variables may be independent of one another, in which case one could not function as a mediator of the age-related effects on the other.

Analytic Strategy

The questions of convergent and discriminant validity and the role of executive functioning as a potential mediator of age–cognition relations were investigated with correlation-based pro-

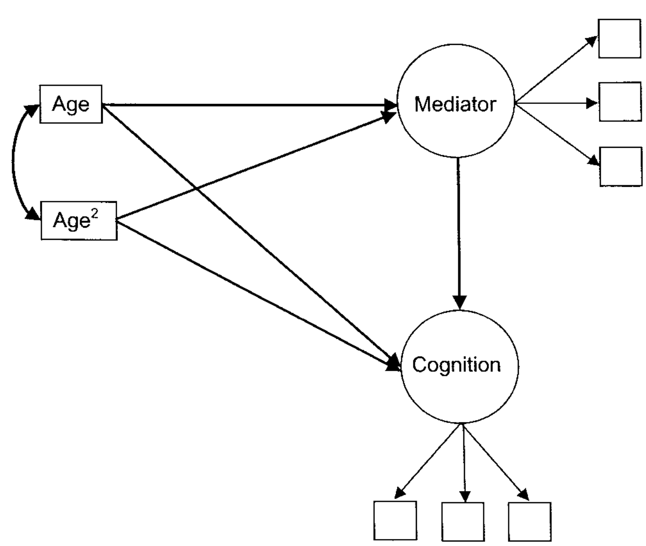

Figure 3. Model representing the mediation of both linear (age) and quadratic (age^2) age-related influences on a cognitive construct through a mediator construct. The boxes represent measured variables, and the circles represent latent constructs.

cedures. Specifically, confirmatory factor analyses and structural equation modeling were used to investigate the relations portrayed in Figures 2 and 3. These types of analytical procedures often involve postulating alternative models that represent different hypotheses about the presence or absence of specific relations among the variables or constructs, and then determining which model provides a significantly better fit to the data. Although the model comparison approach provides a statistical test of whether particular parameters are equal to a specified value, not all questions are amenable to all-or-none conclusions (e.g., that a correlation is, or is not, equal to 1.0). Because a primary goal in the current project was to estimate the relative strengths of the paths corresponding to the thick lines (representing convergent validity) and the dotted lines (representing discriminant validity) in Figure 2, an alternative parameter-estimation approach was used in which the magnitudes of different relations were examined within the same basic model. However, three indices of fit were reported for each model, the chi-square test of whether the model perfectly fits the data; the comparative fit index (CFI), indicating the relative improvement over an alternative model with no relations among variables; and the root-mean-square error of approximation (RMSEA), representing the difference between predicted and observed covariances (see Kline, 1998, for a discussion of these indices).

Method

Participants

Descriptive characteristics of the sample are presented in Table 1. Data from an additional nine participants were not included in the analyses because these individuals failed to complete all of the sessions ($n = 6$; ages 18, 20, 25, 26, 56, and 85) or because they seemed to have difficulty understanding the instructions in several of the tasks and had vocabulary scaled scores of less than 4 ($n = 3$; ages 68, 84, and 95). It should be noted that the participants varied considerably in age but that the average level of education was nearly 16 years, corresponding to 4 years of college, and that most participants rated their health in the very good to excellent range.

Because of reports that age differences vary according to time of testing and self-reported time-of-day preference (e.g., Intons-Peterson et al., 1998; May, 1999; May & Hasher, 1998), the starting time for each session was recorded and the participants were asked to complete the Morningness–Eveningness Questionnaire (Horne & Ostberg, 1976). Participants selected their own times to start the testing, and these times ranged from 8 a.m. to 6 p.m. Increased age was associated with a greater self-reported morning preference (i.e., $r = .45$) and was also associated with significantly earlier starting times in each session (i.e., *r*s = –.15, –.18, and –.19, for Sessions 1, 2, and 3, respectively). However, there was a wide distribution of ages within each starting time. To illustrate, in each of the sessions, between 37% and 55% of the 66 or more adults who started testing before 10 a.m. were under the age of 50, and between 31% and 34% of the 88 or more adults who started testing after 3 p.m. were over the age of 50.

An indication of the representativeness of the sample can be obtained by converting scores on the standardized tests to age-adjusted scaled scores (which have a mean of 10 and a standard deviation of 3) based on the normative sample from the Wechsler Adult Intelligence Scale—Third Edition (WAIS–III; Wechsler, 1997a). These values are in the bottom of Table 1, where it can be seen that the sample had an average that was approximately 0.5 to 0.7 standard deviations above the age-adjusted means from the nationally representative samples used for the norms. Correlations of the scaled scores with age were .11 for Vocabulary, .09 for Digit Symbol, .16 for Logical Memory, and .02 for Recall Memory. These results indicate that although the participants performed higher than the

Table 1
Means (and Standard Deviations) of Demographic Characteristics of the Participants

Variable	Age 18–39	Age 40–59	Age 60–84	Total
N	79	112	70	261
Age	27.7 (6.4)	49.0 (5.0)	70.3 (6.2)	48.2 (17.2)
% females	57	71	63	65
Health rating[a]	2.0 (0.7)	2.0 (0.9)	2.1 (0.9)	2.0 (0.8)
Activity limitations[b]	1.3 (0.7)	1.5 (0.8)	1.8 (0.9)	1.5 (0.8)
Number of medications[c]	0.6 (0.9)	1.3 (1.6)	4.0 (10.7)	1.8 (5.8)
Vision—right eye[d]	0.95 (0.25)	0.63 (0.31)	0.53 (0.22)	0.70 (0.32)
Vision—left eye[d]	0.99 (0.25)	0.64 (0.30)	0.57 (0.22)	0.73 (0.32)
Education[e]	15.5 (3.3)	16.0 (2.4)	16.4 (2.9)	15.9 (2.8)
Mrng-Eve[f]	46.3 (10.3)	57.1 (9.8)	58.9 (9.2)	54.4 (11.1)
Median starting times[g]				
Session 1	14:00 (3.5)	13:00 (3.5)	13:00 (3.4)	13:17 (3.2)
Session 2	13:00 (3.3)	13:00 (3.1)	12:00 (3.2)	12.92 (3.2)
Session 3	14:00 (2.3)	12:00 (2.6)	11:50 (2.6)	12:82 (3.2)
Scaled scores[h]				
Vocabulary	12.0 (3.5)	12.2 (2.7)	12.8 (2.4)	12.3 (2.9)
Digit Symbol	11.7 (3.5)	12.0 (3.1)	12.3 (2.4)	12.0 (3.1)
Logical Memory	11.1 (2.8)	11.4 (2.5)	12.2 (2.9)	11.5 (2.7)
Free Recall	12.1 (3.0)	12.9 (2.9)	12.3 (3.0)	12.5 (3.0)

[a] Self-rating on a scale from 1 (*excellent*) to 5 (*poor*). [b] Self-rating on a scale from 1 (*not at all*) to 5 (*a great deal*). [c] Self-report of number of prescription medications taken per week. [d] Visual acuity in Snellen ratios (e.g., 20/40 = 0.50). [e] Self-report of number of years of formal education completed. [f] Scores on the Morningness–Eveningness Questionnaire (Horne & Ostberg, 1976). [g] Times reported in 24-hr time. [h] Age-adjusted scores based on the norms from the Wechsler Adult Intelligence Scale, Third Edition (Vocabulary and Digit Symbol) and Wechsler Memory Scale, Third Edition (Logical Memory and Free Recall). The scaled scores in the normative sample have a mean of 10 and a standard deviation of 3.

national averages, this was true to nearly the same extent for individuals at each age.

All participants came to the laboratory for three sessions of approximately 2 hr each, with most of the participants completing the three sessions within a 2-week interval. The sample was recruited from newspaper advertisements and flyers, and each participant received a compensation of $100.

Cognitive Ability Tests

Fifteen tests were administered to assess cognitive abilities of vocabulary, fluid intelligence (inductive reasoning and spatial visualization), episodic memory, and perceptual speed. Commercially published tests were administered and scored according to the published manuals. The sources of the tests, and a brief description of each, are as follows.

Vocabulary

WAIS–III Vocabulary (Wechsler, 1997a). This test required the participant to provide a definition of a word spoken by the examiner.

Woodcock–Johnson Revised Picture Vocabulary (Woodcock & Johnson, 1990). In this test the participant attempted to name pictured objects.

Antonym vocabulary (from Salthouse, 1993). In this test the participant was to select the correct antonym from among five alternatives.

Synonym vocabulary (from Salthouse, 1993). This test was similar to the antonym vocabulary test except that here the participant was to select the correct synonym from among five alternatives.

Fluid Intelligence (Reasoning and Spatial Visualization)

Matrix reasoning (Raven, 1962). In this test the participant attempted to determine which pattern best completed the missing cell of a matrix. Three sample problems were followed by the 18 odd-numbered items from the Advanced Progressive Matrices. The participant was allowed 10 min to work on the 18 test problems.

Letter sets (Ekstrom, French, Harman, & Dermen, 1976). Ten minutes were allowed for the participant to select which group of letters did not belong in each of 20 sets of letters.

Spatial visualization (Bennett, Seashore, & Wesman, 1997). In this test the participant was to select which of four three-dimensional objects corresponded to an unfolded two-dimensional drawing. Ten minutes were allowed to solve 20 problems.

Paper folding (Ekstrom et al., 1976). In this test the participant attempted to determine which of five patterns of holes would result from the displayed sequence of folds and hole punches. There were 12 problems, and the time limit in the test was 10 min.

Form boards (Ekstrom et al., 1976). Eight minutes were allowed to solve 24 problems consisting of determining which combination of five shapes was needed to create a complex figure.

Episodic Memory

Wechsler Memory Scale—Third Edition (WMS–III) Logical Memory (Wechsler, 1997b). In this test the examiner read a story (of 65 words), and the participant attempted to recall as much of it as he or she could remember. A second story (of 86 words) followed immediately after the first, and then was repeated with another recall attempt.

WMS–III Word Recall (Wechsler, 1997b). This test involved the auditory presentation of a list of 12 unrelated words four times, with the participant attempting to remember as many words as possible after each presentation. A different list of words (List B) was then presented and recalled, followed by an attempt to recall as many of the words from the original list.

Paired associates (Salthouse & Miles, 2002). This task involved the auditory presentation of two trials of six word pairs each, with the participant instructed to recall the second member of the pair when presented with the first.

Perceptual Speed

WAIS–III Digit Symbol (Wechsler, 1997a). Two minutes were allowed for the participant to write symbols below digits according to a code table displayed at the top of the page.

Letter comparison (Salthouse & Babcock, 1991). Thirty seconds were allowed on each of two pages for the participant to write the letter *S* for same or *D* for different between 21 pairs of 3, 6, or 9 letters.

Pattern comparison (Salthouse & Babcock, 1991). Thirty seconds were allowed on each of two pages for the participant to write the letter *S* for same or *D* for different between 30 pairs of patterns consisting of 3, 6, or 9 line segments.

Neuropsychological Tests

WCST. The computer-administered version of this task described in Fristoe et al. (1997) was used. The task involved the display of four “cards” containing one, two, three, or four colored shapes. The participant was instructed to sort each new card below one of the four existing cards by pressing the appropriate number on the keyboard (i.e., 1 to 4). No specific instructions regarding the basis for the sorting were provided, but feedback in the form of the words *right* or *wrong* was presented after every response. The dimension (i.e., color, form, or number) used to determine the “correct” sorting changed after 10 correct sorts, and the sequence terminated after six such changes (i.e., “categories”) or on completion of 128 trials (i.e., “cards”). A variety of different dependent variables are available in this task, with the most commonly used consisting of the number of categories correctly solved and the number of perseverative errors.

Towers (from the Multitowers program developed by Hansjorg Neth, 2002). Stimuli in this computer-administered task consisted of two displays of three towers and either two, three, or four disks of different sizes and colors. The top display was the goal state and the bottom was the starting state. The task for the participant was to transform the bottom configuration to match the top configuration by moving one disk at a time (by clicking above the tower with a mouse), and never placing a larger disk on top of a smaller disk. Instructions, presented at several periods during the performance of the task, emphasized that it was important to solve the problems with the fewest moves possible.

Two sample problems were presented followed by six test problems. The problems can be described in terms of the start states and the goal states, with the disks from smallest to largest designated A through D and the towers from left to right designated 1 through 3. For example, A1, B2 would indicate that the smallest disk was on the left tower and the next smallest disk was on the middle tower. Using this terminology the start states and goal states for the problems were as follows: Sample 1: A3, B1 to A1, B3; Sample 2: A2, B3 to A2, B2; Problem 1: A1, B2 to A2, B1; Problem 2: A2, B2 to A3, B1; Problem 3: A2, B2 to A3, B3; Problem 4: A1, B1, C1 to A3, B3, C3; Problem 5: A2, B2, C2 to A3, B3, C3; and Problem 6: A2, B2, C2, D2 to A3, B3, C3, D3. The dependent variable was the number of moves required to achieve the correct goal state for each problem.

Connections (from Salthouse et al., 2000). This variant of a trail-making task involves the participant connecting 49 numbers, letters, or intermixed numbers and letters in sequence as rapidly as possible. Each condition was presented on a separate page in the order numbers, numbers–letters, letters–numbers, and letters, followed by the same conditions (with a different arrangement of items) in reverse order. Twenty seconds were allowed for each page, and the number of correct connections was divided into 20 to yield a measure of time per connection as the dependent variable.

Verbal fluency (Delis et al., 2001). This task, sometimes known as controlled associations, involves the participant writing as many words as possible that begin with a designated letter (i.e., F, A, and S on separate trials) in 60 s. The instructions stated that proper names or variants of the same word with different suffixes were not to be produced. The primary dependent variable was the number of unique responses, but the number of clusters and the average words per cluster were also computed.

Figural fluency (Ruff, 1996). In this test the participant was instructed to draw lines connecting dots to create as many unique designs as possible. Each of five pages contained different configurations of dots, and 60 s were allowed per page. The primary dependent variable was the number of unique designs, but the number of repeated designs was also recorded.

Experimental Tasks

Stroop (from Salthouse & Meinz, 1995). Stimuli in this task consisted of pages of 20 letter strings, each in one of four colors (red, blue, green, and yellow). All of the letter strings (i.e., the words *red*, *blue*, *green*, and *yellow*, or a string of 3 to 6 *X*s) on a single page were from the same condition, with the three conditions consisting of congruent (color same as word), neutral (*X*s), and incongruent (color different from word). In each case the task was to name all of the ink colors on the page as rapidly as possible. The conditions were presented in the order C (congruent), N (neutral), I (incongruent), C, N, I, C, N, I, C, N, and I. The dependent variable was the time (in seconds) to name all of the colors on the page.

Reading with distraction (from Connelly, Hasher, & Zacks, 1991). Stimulus material in this task consisted of four short passages with the relevant material printed in italic type (in 10-point Courier font). In the first ("A Day at Work") and the last ("At the Liquor Store") passages, all of the material was relevant and was presented in italic font. In the second ("A New Home") and third ("At the Eye Doctor's") passages, words printed in normal type (10-point Courier font) were interpolated within the passage. The task was to read the relevant story as quickly as possible consistent with good comprehension and to ignore the distracting material. Each story was followed by four comprehension questions, with each question containing six alternative answers. The dependent variables were the time to read the relevant story and the percentage of correct responses for the comprehension questions.

Parameters for the computer-administered tasks were selected on the basis of several pilot studies to maximize sensitivity and reliability. All of the tasks were programmed with E-Prime (2002), and the stimuli were presented on 19-in. flat panel displays. The reported visual angles are based on a viewing distance of 50 cm, but because head position was not constrained the values are only approximate.

Anti-cue (similar to de Jong, 2001). Stimuli in this task were the letters *P* and *Q*, and the responses consisted of presses of those keys. The "+" symbol was used as a fixation and appeared for 500 ms between each trial. The target stimuli were presented for 75 ms, but 350 ms before each stimulus an asterisk flashed (i.e., 50 ms on, 25 ms off, and 50 ms on) either on the same side as the target stimulus or on the opposite side. Participants were instructed to try to ignore this flash. The letter stimuli were each approximately 0.9° high and were displayed approximately 12.5° to the left or right of the center of the screen. A visual mask (the "#" symbol) was presented immediately after the stimulus in the location of the possible stimulus on both sides of the display and remained until a response was registered. Practice before the experimental trials consisted of six trials without a poststimulus mask and six trials with the mask. The experimental trials were presented in four blocks of 48 trials each. Dependent variables were the median reaction times[2] for both correct and incorrect trials and the percentage of correct responses for congruent (flash and target on the same side) and incongruent (flash and target on different sides) trials.

Stop signal (after Logan, 1994). Stimuli in this task were the letters *X* (to which the response was a press of the *Z* key) and *O* (to which the response was a press of the *M* key). Each trial began with a 1,000-ms fixation (the "+" symbol) followed by a 50-ms target stimulus. On 25% of the trials an auditory signal (1000 Hz for 100 ms) occurred, with equal probability at 200, 400, or 600 ms after the target stimulus. The auditory signal indicated that the participant was to refrain from making any response to the previous target stimulus. The letters each subtended a visual angle of approximately 1.3°. There were six practice trials followed by four blocks of 48 trials each. Dependent variables were the median reaction time for both correct and incorrect trials and percentage of correct responses in the *X-O* decision, and the probability of not making the response in the stop-signal trials at each delay interval.

Keeping track (after Yntema & Mueser, 1960). In this task a block of trials began with an initial display of the names of the four categories to be used in the block. A sequence of exemplars from the categories was then presented, followed by category-exemplar probes for which the participant was instructed to press the *Z* key if the exemplar was the last-presented member of the category and to press the *M* key if it was not the last-presented exemplar. All of the exemplar probes were from the designated category, but 50% consisted of the most recently presented exemplar and 50% consisted of an exemplar presented earlier on that trial. There was a 300-ms blank screen between items, and the items were displayed for 1,000 ms, except for the category-exemplar probes, which remained visible until terminated by the registration of a response. The approximate size of the letters in the items was about 0.70°. Four practice trials were followed by two blocks of 16 trials each, with from 0 to 7 items intervening between the last exemplar and the probe. The categories in the practice block were birds, metals, rooms, and vegetables. Categories for the first block were jewelry, fruits, furniture, and seasons, and categories for the second block were animals, directions, shapes, and tools. The primary dependent variable was the percentage of correct decisions, but accuracy was also computed for probes with short lags (i.e., 0 to 3 intervening items) and for probes with long lags (i.e., 4 to 7 intervening items).

Digit monitoring (locally developed). Participants in this task were to respond to every third odd digit by pressing the letter *Z* and to respond to all other items by pressing the letter *M*. The digits (0 through 9, each approximately 1.6° in size) were displayed until a response was registered, after which a 200-ms delay occurred before presentation of the next stimulus. Five practice trials were followed by two blocks, each containing 280 digits and 28 targets. The primary dependent variable was the percentage of correct responses, but hit rates, false alarm rates, and the difference between hit rate and false alarm rate were also recorded.

Matrix monitoring (locally developed). Stimuli in this task consisted of a 4 × 4 matrix approximately 11° wide and cells about 2.2° on each side. A target dot was centered in one of the cells and had a diameter of approximately 2°. The matrix with the dot was displayed for 2,500 ms, and then two arrows pointing either left, right, up, or down were successively presented for 1,000 ms each with an interstimulus interval of 250 ms. These arrows indicated that the target dot had moved a distance of one cell in the direction of the arrow. The probe display, consisting of a dot in one cell of the matrix, was displayed until a response (*Z* for "yes the dot is in the correct position" and *M* for "no the dot is not in the position" corresponding to where it should be according to the initial position and sequence of arrows) was registered. Two versions of the task were administered, one in which only a single matrix was presented on each trial and one in which two matrices were displayed with independently determined cell locations and arrow directions. Only one of the two matrices was probed in the two-matrix condition, each with equal probability, and in both conditions the dependent variable was the percentage of correct responses. Four practice trials were followed by two blocks of 16 trials each with one matrix, and then two blocks of 16 trials each with two matrices.

Tracking with paired associates (Salthouse & Miles, 2002). In this activity two tasks were performed alone and concurrently. The visual–motor tracking task involved the participant using a track ball to try to keep a cursor on a randomly moving white circle. The difficulty of the tracking

[2] Medians were used in the analyses because they are less sensitive to extreme observations that have been postulated to increase with age (e.g., Bunce, Warr, & Cochrane, 1993). However, mean reaction times were also computed and had correlations with the medians of .93 in the congruent condition and .95 in the incongruent condition.

task was first adjusted for each research participant by altering the average magnitude of the shift in target position to achieve a root-mean-square error (RMSE) between cursor and target position of between 40 and 50 pixels (corresponding to approximately 1.0° of visual angle). That is, the average change in position was increased if the individual's error in the preceding trial was less than 40, and it was decreased if the error was greater than 50. The adjustment trials, which were each 20 s in duration, were repeated until the RMSE was within the target range. All subsequent tracking trials were presented at the individually determined level of difficulty for 70 s, and the dependent variable was the RMSE on each trial.

Following the adjustment phase, the single- and dual-task conditions were administered with one trial of tracking alone, one trial of the paired-associates task alone, six trials with the paired-associates task and the tracking task performed together, one trial of the paired-associates task alone, and one trial of tracking alone. In the dual-task conditions, the paired-associates task started 5 s after the tracking task began and finished 5 s before the tracking task ended. Stimuli in the paired-associates task were presented through a pair of speakers, and the participant's vocal responses were recorded by the examiner. Participants were instructed to emphasize performance on the paired-associates task but also to perform the tracking task as well as possible given that constraint.

The paired-associates task involved the auditory presentation of six pairs of words with a 4-s interval between successive pairs, followed by presentation of the first member of each pair at 5-s intervals. All words were concrete nouns of at least moderate frequency. The dependent variable was the number of response words correctly recalled to the appropriate stimulus words.

Driving with n*-back (locally developed).* In this activity a simulated driving task was performed alone and together with the *n*-back task. In the driving task the participant used a steering wheel to try to keep a cursor, representing the vehicle, between curved parallel lines, representing the road. An initial 20-s trial was performed to familiarize the participant with the apparatus, and then two 20-s trials were administered with the speed of the cursor adjusted every second according to the magnitude of error. The average speed across the last 5 s of the two trials was used in all subsequent trials.

Stimuli in the *n*-back task consisted of a sequence of 25 digits presented at a rate of one every 2.5 s. Participants were instructed to vocally repeat the digits that occurred *n* items back in the sequence, which were recorded by the examiner. If the participant was incorrect or failed to respond, the examiner immediately said, "Wrong, start over," which instructed the participant to resume the sequence as quickly as possible. Participants were instructed to emphasize performance in the *n*-back task but also to perform the driving task as well as possible given that constraint.

The *n*-back task was first presented in the order 0 back, 1 back, and 2 back. The driving task was then performed alone, followed by dual-task driving in the order 0 back, 1 back, 2 back, 2 back, 1 back, and 0 back. Finally driving was performed alone followed by the *n*-back task alone in the order 2 back, 1 back, and 0 back. Each driving trial was 70-s long, and the *n*-back task started 5 s after driving began and ended 5 s before driving ended. The dependent variable in the *n*-back task was the number of errors in repeating the sequence of digits, and in the driving task it was the RMSE between the cursor position and the middle of the "road."

Connect and count (locally developed). The two tasks in this dual-task activity were a connections task and a counting backward task. The connections task was similar to the letters condition of the earlier-performed connections task in which the participant attempted to connect the letters in alphabetical sequence as rapidly as possible. However, in this version the circled letters were distributed haphazardly on the page instead of being adjacent to one another as in the other connections task. The counting task consisted of counting backward by 3s from a number read by the examiner. The dependent measure in the connections task was the time to complete the connections between the 26 circles. Participants were instructed to emphasize performance in the count-back task but also to perform the connections task as well as possible given that constraint. The counting task was administered for 20 s when performed alone, and in the dual condition it continued until completion of the 25 connections. The time per subtraction served as the dependent variable in the counting task.

Results

Across the 30 tasks and 261 participants less than 0.5% of the data were missing, and the largest percentage of missing data for any single task was 2.7%. To simplify subsequent analytical procedures, all missing values were replaced with the mean of the variable.[3]

The initial analyses of the data examined interactions of age with gender, self-reported health, visual acuity, years of education, time of testing, and morningness–eveningness score on the 102 variables reported in Tables 3 through 7 (which will be discussed later in the text). All of the statistically significant ($p < .01$) interactions involving age on these variables are reported in Table 2. Perhaps the most interesting results are those concerning the time-of-testing and morningness–eveningness variables. Unlike earlier studies involving comparisons of college students and older adults (e.g., Intons-Peterson et al., 1998; May, 1999; May & Hasher, 1998), only a few variables had significant interactions, and they were all in the direction opposite of the expectation, as the age-related effects were more negative when the testing was in the morning or when the participants expressed a preference for morning. The triple interaction of Age × Morningness–Eveningness × Time of Testing was not significant for any of the 102 variables.

Because only a few of the interactions of age with gender, health, visual acuity, years of education, time of testing, or morningness–eveningness were statistically significant, and all were small in magnitude, these factors were ignored in all subsequent analyses.

Psychometric Variables

The six recall variables (i.e., recall on the first four trials with the original list, recall on the new list, and recall on the original list again) and the three logical memory variables (i.e., recall of the first story and the two recall attempts of the second story) were initially examined with separate exploratory factor analyses to determine the number of distinct factors represented by these variables. In both cases a single factor accounted for more than 65% of the variance, and therefore the sum of the number of words recalled across Trials 1 through 4 was used as the summary measure of recall performance, and the sum of the three logical memory scores was used as the summary logical memory measure.

A confirmatory factor analysis was then carried out with a five-factor model in which the matrix reasoning and letter sets variables were hypothesized to represent inductive reasoning, and the spatial relations, paper folding, and form boards variables were hypothesized to represent spatial visualization, in addition to variables representing factors corresponding to episodic memory, per-

[3] Mean substitution is a conservative method of dealing with missing data and may lead to attenuation of correlation coefficients (Schafer & Graham, 2002). However, this method was considered justifiable because of the small proportion of missing data. It should also be noted that the parameter estimates were very similar when the analyses were repeated using the full-information maximum likelihood algorithm to estimate missing values within the AMOS program (J. L. Arbuckle & Wothke, 1999).

Table 2
Statistically Significant Interactions With Age

Variable	R^2	Larger negative (or smaller positive) age effect
Age × Gender		
Average *SD* (paired associates & tracking)	.027	Females
Age × Health		
n-back 0-back alone	.043	Poor health
Anti-cue I-C RT difference	.027	Poor health
Age × Visual Acuity		
n-back 0-back alone	.028	Poor vision
Age × Education		
Digit symbol	.025	Higher education
Letter comparison	.021	Higher education
Picture vocabulary	.024	Higher education
Age × Time of Testing		
Synonym vocabulary	.032	Morning
Figural fluency	.020	Morning
Connections alternating sequence	.026	Morning
Age × Morningness–Eveningness		
Digit symbol	.023	a.m. preference
Picture vocabulary	.030	a.m. preference
Paired associates	.040	a.m. preference
Connections same	.020	a.m. preference
Connections alternating	.027	a.m. preference

Note. Interactions are significant at $p < .01$. I-C = incongruent–congruent; RT = reaction time.

ceptual speed, and vocabulary. However, the correlation between the reasoning and spatial factors in this analysis was quite high (i.e., .90), and thus the two reasoning and three space variables were combined to represent a single fluid intelligence (gF) construct.

The four-factor model consisting of gF, memory, speed, and vocabulary abilities was examined with a confirmatory factor analysis and found to yield an acceptable fit to the data, $\chi^2(84, N = 261) = 239.28$, CFI = .987, RMSEA = .084. The matrix of standardized residuals was examined, but no theoretically justifiable relations were found that would improve the fit of the model. Loadings of the variables on the factors and correlations among the factors are presented in Table 3. Also in this table are the estimated reliabilities of the variables and correlations with age. It can be seen that all but one of the reliability estimates were greater than .75 and all of the age correlations were significantly different from zero. The loadings of the variables on the factors ranged from .65 to .90, and all factors were significantly correlated with one another, with values ranging from .18 to .69.

As has been reported many times in the past (e.g., Salthouse, 1998, 2001; Salthouse & Czaja, 2000; Salthouse, Fristoe, & Rhee, 1996; Salthouse, Hambrick, & McGuthry, 1998), there were significant negative correlations between age and the gF, memory, and speed factors and a significant positive correlation between age and the vocabulary factor. To illustrate the age trends on the cognitive abilities, the variables loading on each factor were converted into *z* scores and then averaged to form composites. Means and standard errors of these composite scores by age decade are displayed in Figure 4.

The apparent nonlinearity in the age trends in Figure 4 was examined in hierarchical regression analyses by successively entering the linear age term, the quadratic age term, and finally the cubic age term. None of the cubic effects were significant, but all of the linear and quadratic effects were significantly different from zero. The proportions of variance associated with the linear and quadratic trends, were, respectively, .113 and .056 for vocabulary, .181 and .038 for gF, .145 and .034 for memory, and .273 and .022 for speed. Inspection of Figure 4 reveals that the quadratic trends were attributable to an acceleration of the negative age relations with increased age for the gF, memory, and speed composites, and a deceleration of the positive age relation with increased age for the vocabulary composite. The nonlinear trends were incorporated into the mediational analyses by including both age and age-squared variables in the models.

Neuropsychological Variables

Summary characteristics for the neuropsychological variables used to assess executive functioning, including means by double decade, estimated reliabilities, and age correlations, are contained in Table 4.

An exploratory factor analysis, with promax rotation and factors identified on the basis of the eigenvalue-greater-than-one criterion, was conducted on the WCST variables to determine the number of distinct factors represented by these variables. Three factors were identified in this analysis. The variables with the largest loadings on the first factor were total errors, number of categories, number of perseverative errors, and number of nonperseverative errors. Variables with the largest loadings on the second factor were number of correct responses, number of conceptual responses, and failure to maintain set, and the variables with the highest loadings on the third factor were the trials-to-first-category and learning-to-learn measures. Only the first factor was significantly correlated with age (i.e., $r = .26$), as the correlation with Factor 2 was –.02 and that with Factor 3 was .03. Because the total errors variable had the highest loading on the first factor and was highly correlated with other variables (i.e., .84 with perseverative errors and

Table 3
Standardized Coefficients From Confirmatory Factor Analysis on Psychometric Variables

	Factor					
Variable	gF	Memory	Speed	Vocabulary	Est. rel.	Age *r**
Matrix reasoning	.81				.76	−.46
Letter sets	.70				.76	−.31
Spatial relations	.83				.87	−.28
Paper folding	.80				.81	−.41
Form boards	.71				.81	−.30
Logical memory		.65			.82	−.15
Recall		.73			.86	−.42
Paired associates		.66			.60	−.35
Digit symbol			.82		—[a]	−.48
Letter comparison			.78		.82	−.37
Pattern comparison			.74		.91	−.49
Vocabulary (WAIS–III)				.85	.89	.20
Picture vocabulary				.82	.88	.32
Synonym vocabulary				.90	.86	.40
Antonym vocabulary				.88	.85	.30
Factor						
gF	—					
Memory	.69	—				
Speed	.66	.64	—			
Vocabulary	.39	.37	.18	—		
Age	−.45*	−.47*	−.58*	.36*		

Note. Est. rel. = estimate of reliability; WAIS–III = Wechsler Adult Intelligence Scale, Third Edition; gF = fluid intelligence.
[a] No reliability estimate was available in this study. All other reliability estimates were based on coefficient alpha or were calculated by using the Spearman–Brown formula to boost the correlation between scores on odd-numbered and even-numbered items.
* $p < .01$.

–.89 with number of categories), it was used as the summary measure of WCST performance in subsequent analyses.

Performance in the Tower of Hanoi task was represented by the average number of moves to convert the starting state to the goal state. Most participants were close to the minimum number of moves in the two-disk problems, and thus reliability was very low. Reliability was also low in the three-disk problems because there was little consistency for a given individual in the number of moves across the two trials. Only one four-disk problem was presented, and thus no measure of reliability was available in this condition. Because it was significantly related to age, the three-disk variable appeared to be the most sensitive, and therefore it was used in subsequent analyses.

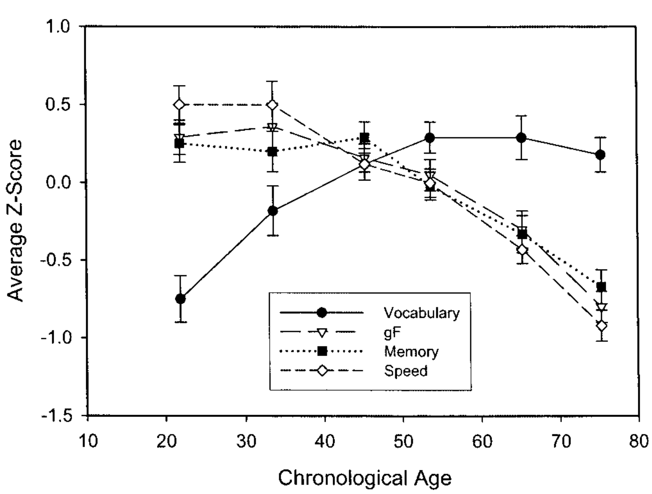

Figure 4. Relation between age and composite scores representing the four cognitive abilities. Data points correspond to the mean and standard error for the 35 to 62 individuals in each decade. gF = fluid intelligence.

Three measures of performance were available in the verbal fluency task: the number of unique words, the number of clusters, and the number of words per cluster, with clusters identified using criteria described by Troyer et al. (1994) and modified by Rende, Ramsberger, and Miyake (2002). Similar negative age relations were evident in the number of unique words and in the number of clusters, but the number-of-words-per-cluster variable was not significantly related to age. Because of its simplicity and reliability, the number of unique words was used to represent verbal fluency performance in subsequent analyses.

Performance in the figural fluency task was assessed by the number of unique designs and by the number of repetitions, in which the participant repeated the same design. Increased age was associated with a smaller number of unique designs and with a slight increase in the number of repeated designs. Because of its simplicity and reliability, the number of unique designs was used as the measure of figural fluency performance in subsequent analyses.

In the connections task increased age was associated with slower performance in both same-sequence (all numbers or all letters) and alternating-sequence (intermixed letters and numbers) conditions. Because it is not obvious whether it is more meaningful to evaluate age-related effects in absolute terms or in relative

Table 4
Means (and Standard Deviations) of Variables From Neuropsychological Executive Functioning Tasks

Task	Age 18–39	Age 40–59	Age 60–84	Est. rel.	Age *r*
Wisconsin Card Sorting Test					
Correct responses	73.9 (12.9)	71.7 (14.7)	67.9 (16.6)	—[a]	−.18*
Categories completed	4.7 (1.8)	4.1 (2.2)	3.2 (2.0)	—[a]	−.30*
Total errors	36.7 (22.5)	41.2 (23.7)	53.7 (22.6)	—[a]	.30*
Perseverative errors	17.4 (12.2)	20.8 (16.5)	25.3 (17.3)	—[a]	.22*
Perseverative responses	19.7 (15.4)	23.8 (21.7)	37.4 (50.9)	—[a]	.21*
Nonperseverative responses	19.9 (13.8)	22.9 (24.3)	26.6 (15.1)	—[a]	.15*
Trials to first category	22.1 (20.5)	20.6 (14.2)	25.1 (20.3)	—[a]	.14*
Conceptual responses	61.5 (20.8)	58.9 (21.6)	53.1 (22.2)	—[a]	−.19*
Failure to maintain set	1.2 (1.2)	1.4 (1.5)	1.6 (1.8)	—[a]	.10
Learning to learn	−0.8 (5.4)	−2.4 (6.0)	−3.2 (6.1)	—[a]	−.10
Tower of Hanoi					
2-disk (average no. moves)	2.8 (0.4)	2.8 (0.3)	2.9 (0.3)	.17[b]	.12
3-disk (average no. moves)	10.4 (2.7)	11.3 (3.4)	12.8 (3.9)	.29[b]	.29*
4-disk (average no. moves)	25.8 (11.2)	29.6 (13.7)	27.9 (11.4)	—[a]	.06
Verbal fluency					
No. unique words	12.9 (3.4)	13.6 (3.3)	11.5 (3.1)	.87[c]	−.15
No. clusters	9.1 (2.6)	9.7 (2.3)	8.2 (2.4)	.76[c]	−.12
Words/cluster	1.5 (0.3)	1.5 (0.3)	1.5 (0.5)	.48[c]	.04
Figural fluency					
No. unique designs	19.5 (5.1)	18.3 (3.5)	14.6 (4.5)	.93[c]	−.41*
No. repetitions	1.3 (1.2)	1.6 (1.5)	1.9 (2.4)	.77[c]	.16*
Connections					
Same sequence[d]	0.66 (0.17)	0.78 (0.21)	1.06 (0.35)	.91[c]	.56*
Alternating sequence[d]	1.39 (0.57)	1.48 (0.52)	2.11 (0.92)	.70[c]	.41*
Difference	0.72 (0.50)	0.70 (0.43)	1.05 (0.74)	.68[c]	.24*
Ratio	2.12 (0.80)	1.91 (0.50)	2.01 (0.68)	.13[c]	−.05

Note. Est. rel. = estimate of reliability.
[a] No reliability estimates were available in this study. [b] Reliability was computed by boosting the correlation between odd-numbered trials and even-numbered trials by the Spearman–Brown formula. [c] Reliability was computed with coefficient alpha with scores from each problem or trial as items. [d] In seconds.
* $p < .01$.

terms, both differences and ratios were examined in comparisons of the two conditions. There was a small increase with age in the difference score, but there was no significant relation between age and the ratio of times in the alternating- and same-sequence conditions. The difference score was used in subsequent analyses because it corresponds most directly to the theoretical process of interest and it had moderate reliability.

Age differences were apparent in many of the variables that have been interpreted as evidence for age-related declines in executive functioning. However, it is premature to infer age differences in an executive functioning construct until its construct validity is established and there is evidence that the variables represent a common dimension of variation that is distinct from that represented by other constructs. Before considering this type of evidence, the variables hypothesized to represent specific aspects of executive control will be described.

Inhibition of Prepotent Responses

Characteristics of the variables hypothesized to reflect inhibition of prepotent responses are summarized in Table 5. In the Stroop task, color naming was fastest in the congruent (color same as word) condition, intermediate in the neutral (colored *X*s) condition, and slowest in the incongruent (color different from word) condition. Furthermore, the correlations with age increased from the congruent to the neutral to the incongruent condition.

The contrast between times in the incongruent and neutral conditions is typically interpreted as representing interference attributable to the conflict between word and color name, and the contrast between times in the neutral and congruent conditions is often interpreted as reflecting facilitation. Interference and facilitation scores can be computed both as differences and as ratios. Increased age was associated with larger differences and larger ratios for the interference comparison, but age was not significantly related to either type of facilitation measure. Because the incongruent minus neutral (I – N) difference score is the most common measure of the amount of interference associated with the conflict between color and word, it was used to summarize Stroop performance in subsequent analyses.[4]

Performance in the reading with distraction task was represented by the time to read the passage and by the proportion of compre-

[4] The ratio of times in the incongruent and neutral conditions was also examined as the measure of Stroop interference in analyses of the four construct validity models. The ratio variable had weaker relations to the speed factor than the difference score, but in all other respects, the pattern was very similar with the two Stroop interference measures.

Table 5
Means (and Standard Deviations) of Variables Hypothesized to Reflect Inhibition of Prepotent Responses

Variable	Age 18–39	Age 40–59	Age 60–84	Est. rel.	Age *r*
Stroop					
Congruent[a]	9.2 (2.7)	9.7 (2.7)	10.7 (3.2)	.88	.23*
Neutral[a]	10.2 (2.7)	10.9 (2.3)	12.4 (2.6)	.90	.36*
Incongruent[a]	15.7 (3.8)	18.3 (4.8)	22.0 (5.5)	.92	.51*
I − N diff.	5.5 (2.3)	7.4 (3.7)	9.6 (3.9)	.77	.46*
I/N ratio	1.56 (0.22)	1.69 (0.31)	1.78 (0.28)	.76	.34*
N − C diff.	1.0 (2.2)	1.2 (2.2)	1.7 (2.3)	.59	.13
N/C ratio	1.14 (0.22)	1.16 (0.23)	1.20 (0.24)	.70	.09
Reading with distraction					
Normal time[a]	40.3 (10.7)	42.0 (8.6)	44.9 (6.7)	.93	.19*
Distraction time[a]	120.3 (47.8)	118.5 (52.4)	142.1 (53.8)	.89	.18*
D − Norm. time diff.	79.9 (40.0)	76.6 (49.2)	97.2 (52.5)	.81	.16*
D − Norm. diff. in normal *SD*s	8.9 (4.4)	8.5 (5.5)	10.8 (5.8)	NA	.16*
D/Norm. time ratio	2.95 (0.72)	2.83 (1.09)	3.19 (1.19)	.85	.11
Normal comprehension	.88 (.10)	.85 (.10)	.83 (.13)	.02	−.23*
Distraction comprehension	.71 (.18)	.70 (.17)	.63 (.21)	.45	−.20*
Norm. − D comprehension diff.	.18 (.18)	.15 (.20)	.20 (.24)	.33	.05
Norm. − D diff. in normal *SD*s	1.6 (1.7)	1.3 (1.8)	1.8 (2.2)	NA	.05
Average *SD* diff.	5.2 (2.6)	4.9 (3.1)	6.3 (3.0)	NA	.16*
Anti-cue					
Congruent accuracy (% correct)	84.6 (12.5)	82.6 (8.6)	75.1 (13.3)	.82	−.35*
Incongruent accuracy (% correct)	82.4 (13.7)	77.6 (11.0)	67.9 (10.8)	.84	−.47*
C − I accuracy diff.	2.2 (7.5)	5.0 (11.6)	7.2 (12.9)	.71	.18*
C − I diff. in congruent *SD*s	0.19 (0.63)	0.41 (0.97)	0.60 (1.07)	NA	.18*
Congruent RT (ms)	834 (202)	948 (268)	1190 (361)	.81	.45*
Incongruent RT (ms)	834 (226)	987 (294)	1328 (528)	.84	.48*
I − C RT diff.	0 (86)	39 (198)	138 (333)	.56	.24*
I − C diff. in congruent *SD*s	−0.00 (0.28)	0.12 (0.64)	0.44 (1.07)	NA	.25*
I/C RT ratio	1.00 (0.10)	1.05 (0.19)	1.11 (0.23)	.56	.25*
Average *SD* diff.	.09 (.42)	.27 (.71)	.52 (.89)	NA	.24*
Stop signal					
RT (ms)	720 (158)	726 (168)	732 (158)	.89	.05
Accuracy (% correct)	93.9 (7.9)	95.3 (5.9)	93.3 (8.6)	.79	−.04
StopProb @ 200 ms	.79 (.33)	.76 (.36)	.71 (.36)	.93	−.09
StopProb @ 400 ms	.74 (.34)	.72 (.36)	.67 (.36)	.92	−.07
StopProb @ 600 ms	.63 (.31)	.61 (.36)	.59 (.35)	.90	−.04

Note. All reliabilities were estimated by boosting the correlation between values from odd-numbered and even-numbered blocks (or trials) by the Spearman–Brown formula. NA means that a reliability estimate (est. rel.) was not available. I = incongruent; N = neutral; C = congruent; diff. = difference; D = distraction; Norm. = normal; RT = reaction time; StopProb = probability of stopping a response.
[a] In seconds.
* $p < .01$.

hension questions answered correctly. As expected, reading time was longer and comprehension accuracy was lower when the passages contained distracting material. Increased age was associated with slower reading and lower comprehension in both types of passages and with a larger difference between the reading times in the two conditions.

Because the distraction manipulation had effects on both the reading time and reading comprehension measures, the total effect of distraction would be underestimated if the variables were considered separately. One method of allowing the two variables to be analyzed simultaneously involves converting the reading time differences and the comprehension accuracy differences into *z* scores of the distribution of scores in the normal reading condition and then creating a composite by averaging the two *z* scores. This method is not ideal because it assumes that all participants have the same, equal emphasis on time and accuracy, but it has the important advantage of allowing variables in different units of measurement to be analyzed simultaneously. This measure of the average distraction costs, which was positively correlated with age, was used to summarize performance in the reading with distraction task in subsequent analyses.

Performance in the anti-cue task was represented by median reaction time and decision accuracy to the *P* or *Q* stimuli in the congruent (flash and target on the same side) and incongruent (flash and target on opposite sides) conditions. Consistent with the assumption that the flash induces a reflexive shift of attention, reaction time was slower and decision accuracy was lower when the flash and target stimulus were on opposite sides of the display than when they were on the same side. Increased age was associated with longer reaction times and lower accuracies in each condition, and with larger differences between the two conditions. To consider the time and accuracy measures together, the difference scores in each variable were converted into standard deviations of the congruent condition and then averaged to represent the overall effect of having the cue appear on the opposite side of the display. This average anti-cue effect, which was larger with increased age, served as the summary measure of performance in this task in subsequent analyses.

There were two components in the stop signal task: reaction time and accuracy to the *X* or *O* stimuli, and the probability of not making a response on trials when the auditory stop signal was presented. Accuracy in responding to the *X* or *O* stimuli averaged 94%, and the reaction time averaged about 725 ms, but neither variable was significantly related to age. The absence of a relation between age and reaction time is unusual, and together with the relatively long reaction times for a two-alternative choice reaction time task, it raises the possibility that some participants may have ignored the instructions to respond as rapidly as possible and waited for the stop signal on at least some trials.

From the current perspective the primary interest in the stop signal task concerns the probability of stopping the response to the *X* or *O* stimulus as a function of the interval between the target stimulus and the auditory stop signal. As expected if the response preparation is closer to completion and hence more difficult to inhibit, the probability of stopping the response decreased as the interval between stimulus and stop signal increased. Because the stopping probabilities with the 400-ms interval were in the middle of the range of values, and because the probabilities were all highly related to one another (i.e., correlations were greater than .86), the stopping probability with the 400-ms interval was used as the summary measure of performance in this task in subsequent analyses.

The preceding discussion and the entries in Table 5 indicate that statistically significant age differences are apparent in many of the measures used to assess inhibition of prepotent responses. Although these results could lead to a conclusion that increased age is associated with a decrease in effectiveness of inhibiting prepotent responses, we again suggest that such a conclusion is premature. Before reaching such a conclusion it is important to investigate construct validity of the hypothesized inhibition construct by determining the amount of variance these variables have in common when relations of the variables to other constructs are considered.

Updating

Characteristics of the variables hypothesized to reflect updating are summarized in Table 6. Increased age was associated with a decrease in the percentage of correct responses in the keeping track task, suggesting a decline in the accuracy of updating the status of continuously changing information. The data were further analyzed by comparing accuracy in trials with a short (0- to 3-item) lag between the last exemplar and the probe with accuracy on trials with a long (4- to 7-item) lag. As expected, accuracy was lower with a longer lag, and the age differences were also slightly larger with a long lag than with a short lag. Because the overall percentage correct variable had the highest reliability, it was used as the summary measure of keeping track performance in subsequent analyses.

Performance in the digit-monitoring task was analyzed in terms of both the percentage of correct decisions and hit rate and false alarm rate measures from signal-detection theory. The only measure with a significant relation to age was false alarm rate, but the absolute level of false alarms was very low. Primarily because of its simplicity, the percentage correct measure was used to represent digit-monitoring performance in subsequent analyses.

Two conditions were administered in the matrix-monitoring task, one with a single matrix and sequence of arrows and another with two matrices, each with its own corresponding sequence of arrows. Not surprisingly, accuracy was considerably lower in the two-matrix condition than in the one-matrix condition, and the age differences were also somewhat larger in the two-matrix condition. Because there was less likelihood of ceiling effects on performance in the two-matrix condition, the percentage of correct decisions in this condition was used as the measure of matrix-monitoring performance in subsequent analyses.

Table 6
Means (and Standard Deviations) of Variables Hypothesized to Reflect Updating

	Age				
Variable	18–39	40–59	60–84	Est. rel.	Age *r*
Keeping track					
% correct	71.7 (12.9)	69.6 (9.9)	64.1 (11.7)	.59	−.23*
Lag 0–3 (% correct)	77.2 (15.6)	77.4 (12.8)	72.7 (16.1)	.50	−.13
Lag 4–7 (% correct)	66.2 (14.5)	61.7 (12.9)	56.9 (12.3)	.27	−.23*
Digit monitoring					
% correct	93.6 (6.0)	94.2 (6.9)	92.3 (8.2)	.74	−.04
Hit rate (HR)	.88 (.12)	.89 (.13)	.86 (.16)	.82	−.02
False alarm rate (FA)	.00 (.01)	.01 (.01)	.01 (.02)	.45	.20*
HR − FA	.88 (.12)	.89 (.14)	.85 (.17)	.69	−.04
Matrix monitoring					
One matrix (% correct)	96.6 (11.3)	95.3 (12.0)	91.6 (14.8)	.84	−.16*
Two matrices (% correct)	76.4 (13.5)	76.4 (11.3)	67.6 (13.0)	.71	−.26*
n-back (alone)					
0-back (no. errors)	0.0 (0.0)	0.0 (0.1)	0.3 (1.1)	.02	.24*
1-back (no. errors)	0.8 (1.7)	1.7 (2.7)	3.7 (4.1)	.62	.39*
2-back (no. errors)	3.2 (2.5)	4.2 (2.5)	5.8 (2.6)	.77	.38*

Note. All reliabilities were estimated by boosting the correlation between values from odd-numbered and even-numbered blocks (or trials) by the Spearman–Brown formula. Est. rel. = estimate of reliability.
* $p < .01$.

Three conditions were presented in the *n*-back alone task (i.e., without concurrent driving): one in which the participant was to repeat the just-presented digit (0 back), one in which he or she was to repeat the digit that occurred previously (1 back), and one in which the digit that occurred two earlier in the sequence was to be repeated (2 back). As expected, the number of errors increased from the 0-back to the 1-back to the 2-back conditions, and the age differences were larger in the 1-back and 2-back conditions than in the 0-back condition. No errors were expected in the 0-back condition because it did not require any updating, and the few that did occur may have been attributable to occasional difficulties in hearing the stimuli. Because performance in the 2-back condition was sensitive and reliable, it served as the measure of *n*-back performance in subsequent analyses.

As was the case with the variables hypothesized to represent executive functioning and inhibition, increased age was associated with lower levels of performance on many of the variables hypothesized to reflect effectiveness of updating. However, because age differences are found in a great number of cognitive variables, it is important to investigate whether these variables form a coherent construct before concluding that there are age-related declines in updating aspects of executive control.

Time Sharing

Characteristics of the variables used to assess time sharing are summarized in Table 7. The table contains not only information about the variables in each task when performed alone and when performed with another task but also information about derived variables designed to allow performance in both tasks to be analyzed simultaneously. The method used to combine the variables was proposed by Salthouse and Miles (2002) and is similar to that used in the reading with distraction and anti-cue tasks described above. In the current context it consisted of dividing the difference between performance in the alone and dual conditions by the standard deviation from the alone condition, and then averaging the measures from the two tasks to create an overall measure of time-sharing cost. Because it incorporates effects in both tasks, this average standard deviation difference was used as the measure of time sharing with each combination of tasks in subsequent analyses.

Performance in the paired-associates task was assessed by the number of response terms correctly supplied to the appropriate stimulus term. Increased age was associated with lower accuracy in both alone and dual conditions and with larger age differences when the paired-associates task was performed concurrently with the tracking task.

Before beginning the paired-associates and tracking tasks, the difficulty of the tracking task was adjusted for each participant by varying the size of the difference between successive positions of the target. As a consequence of this adjustment, the position shift used in the actual trials was smaller for older participants (i.e., age correlation = –.35). The adjustment was moderately successful because there was not a significant relation between age and tracking performance in the alone condition. The average paired-

Table 7
Means (and Standard Deviations) of Variables Hypothesized to Reflect Time Sharing

Variable	Age 18–39	Age 40–59	Age 60–84	Est. rel.	Age *r*
Tracking with paired associates (PA)					
PA alone (no. correct)	3.0 (1.5)	3.0 (1.6)	1.8 (1.2)	.60	−.35*
PA dual (no. correct)	3.3 (1.4)	2.9 (1.5)	1.6 (1.0)	.91	−.49*
PA diff. (A − D) in alone *SD*s	−0.22 (0.65)	0.10 (0.71)	0.16 (0.61)	NA	.18*
Track alone (RMSE)	46.4 (7.7)	44.9 (5.5)	48.2 (9.6)	.88	.12
Track dual (RMSE)	48.3 (12.7)	46.8 (7.8)	51.8 (14.1)	.97	.16
Track diff. (D − A) in alone *SD*s	0.24 (0.79)	0.25 (0.60)	0.48 (1.00)	NA	.15
Average *SD*	.01 (.49)	.18 (.47)	.32 (.58)	NA	.23*
Driving with *n*-back 2					
n-back 2 alone (no. errors)	3.2 (2.5)	4.2 (2.5)	5.8 (2.6)	.77	.38*
n-back 2 dual (no. errors)	5.7 (4.3)	7.9 (4.5)	9.9 (5.1)	.84	.37*
n-back 2 diff. (D − A) in alone *SD*s	0.94 (1.17)	1.36 (1.22)	1.53 (1.51)	NA	.22*
Drive alone (RMSE)	25.6 (3.8)	25.6 (3.2)	25.9 (3.7)	.72	.08
Drive dual (RMSE)	26.9 (4.9)	26.6 (4.7)	27.6 (6.4)	.90	.11
Drive diff. (D − A) in alone *SD*s	0.35 (0.86)	0.28 (0.98)	0.49 (1.24)	NA	.08
Average *SD*	0.64 (0.74)	0.82 (0.70)	1.01 (1.07)	NA	.21*
Connect and count					
Connect alone[a]	11.3 (3.1)	12.6 (2.9)	16.4 (4.9)	.94	.51*
Connect dual[a]	62.9 (31.2)	59.8 (22.2)	70.5 (25.9)	.94	.15
Connect diff. (D − A) in alone *SD*s	12.6 (7.2)	11.5 (5.1)	13.2 (5.7)	NA	.08
Count alone[b]	2.3 (1.2)	2.2 (1.0)	2.6 (1.2)	.95	.12
Count dual[b]	4.7 (2.3)	4.7 (2.0)	5.7 (2.5)	.93	.18*
Count diff. (D − A) in alone *SD*s	2.1 (1.3)	2.2 (1.3)	2.7 (1.6)	NA	.18*
Average *SD*	7.4 (3.9)	6.9 (2.8)	8.0 (3.2)	NA	.11

Note. All reliabilities were estimated by boosting the correlation between values from odd-numbered and even-numbered blocks (or trials) by the Spearman–Brown formula. NA means that a reliability estimate was not available. Est. rel. = estimate of reliability; diff. = difference; A = alone; D = dual; RMSE = root-mean-square error.
[a] In seconds. [b] Seconds/number.
* $p < .01$.

associate and tracking time-sharing cost expressed in standard deviations of the alone conditions was significantly greater with increased age.

Only the 2-back condition in the *n*-back task is considered in the current analyses because it was the most sensitive of the *n*-back variables. The number of errors increased when the *n*-back task was performed concurrently with the simulated driving task, but the age relationship was similar in both the alone and dual conditions.

As with the tracking task, difficulty of the driving task was initially adjusted to minimize individual differences in single task performance. The adjusted driving speed was lower with increased age (i.e., $r = -.38$), and there was no significant relation between age and driving error in the alone condition. Driving error increased only slightly when the driving was performed with the *n*-back task, but there was a significant increase with age in the average time-sharing cost.

Performance in the connections task was assessed by the time to draw lines between the 26 letters. This time was considerably longer when the counting backward task was performed concurrently. Surprisingly, the age correlation was smaller in the dual condition than in the alone condition, possibly because of a substantial increase in between-person variability.

Count-back performance was assessed in terms of the time per subtraction, because a fixed duration was used in the alone condition, whereas in the dual condition the subtraction continued until completion of the connections task. Count-back time was much larger when the connections task was performed concurrently, and there was a slight increase in the age correlation. The average time-sharing cost in the connect and count task was not significantly related to age.

The average time-sharing cost had a significant positive correlation with age in two of the three dual-task situations. However, as with the other proposed constructs, it is important to consider construct validity before reaching a conclusion that aging is associated with a decrease in time-sharing aspects of executive control.

Construct Validity

For each hypothesized construct, four structural equation analyses based on the models portrayed in Figure 2 were conducted. The initial analysis included only the target variables and the hypothesized construct (as represented in Panel A of Figure 2). The next analysis allowed that construct to be related to other constructs (as portrayed in Panel B). The next allowed the target variables to be related to other constructs if they resulted in an improved fit (as in Panel C), and the final analysis examined variance common to the hypothesized construct when relations of individual variables to other constructs were considered (as in Panel D). In each model convergent validity would be established by moderate to strong loadings on the hypothesized construct (represented by the thick lines in Figure 2), indicating that the variables have variance in common. Evidence for discriminant validity would be manifested by weak relations (represented by the dotted lines in Figure 2) of the hypothesized construct to other constructs (in Models B and C) or of the candidate variables to other constructs (in Models C and D).

Results of the analyses with the neuropsychological variables hypothesized to reflect executive functioning are summarized in Table 8. The fit statistics in the top panel indicate that each of the models provided a reasonable fit to the data. Of particular interest are the values of the standardized coefficients in the models, and these are summarized in the remaining panels of the table.

The second panel indicates that the variables all had moderate, and statistically significant, loadings on an executive functioning construct in Models A and B. The loading of the verbal fluency variable on the executive functioning construct was no longer significant when it was allowed to load on other constructs in Model C. As noted above, a finding of significant loadings on a common construct can be interpreted as evidence for convergent validity. However, in Models B and C the correlations of the hypothesized executive functioning construct with the gF, memory, and speed constructs were quite large and, in the case of the gF construct, very close to 1.0. Because these correlations were larger than the loadings of any of the hypothesized executive functioning variables on the executive functioning construct, the construct can be inferred to lack discriminant validity.

In Model D most of the loadings of the variables on the executive functioning construct were low, and none of them were significantly different from zero. In contrast, all of the variables had significant relations to other constructs. The coefficients in the bottom panel of Table 8 are interesting because they can be interpreted as reflecting which other cognitive abilities contribute to each variable. For example, based on the pattern of coefficients, the verbal fluency variable can be inferred to be closely related to speed and vocabulary abilities, whereas the figural fluency variable can be inferred to be closely related to gF and speed abilities. Because the variables differed in terms of their patterns of relations to other constructs, and because they had little variance in common when these relations were taken into consideration, the results from the Model D analyses suggest that the executive functioning construct based on neuropsychological variables lacked both convergent and discriminant validity.

The same types of analyses were conducted on variables hypothesized to represent different aspects of executive functioning. An initial analysis examined the three hypothesized constructs simultaneously. This model resulted in a reasonably good fit to the data, $\chi^2(24, N = 261) = 38.11$, CFI = .996, RMSEA = .048, but correlations among the constructs were quite high (i.e., inhibition with updating, $r = .71$; inhibition with time-sharing, $r = .94$; and updating with time-sharing, $r = .90$). This pattern implies that there is considerable overlap in the individual-differences variance among the variables hypothesized to represent different aspects of executive control. In fact, the high correlations suggest that a large proportion of the individual differences in these variables could be accounted for in terms of a single higher order factor. Because inclusion of all of the cognitive process variables in the same analysis would result in a loss in the specificity of the hypothesized theoretical processes, the executive control constructs were considered separately in the following analyses.

Results from the analyses of the variables hypothesized to represent inhibition of prepotent responses are summarized in Table 9. The Model A analysis revealed that the stop signal measure did not have a significant loading on the hypothesized inhibition construct, and thus it was deleted from subsequent analyses. The other three variables had at least moderate, and statistically significant, loadings on the inhibition construct in both Models B and C, suggesting that the construct had convergent validity. Furthermore, the correlations of the inhibition construct with the constructs representing other cognitive abilities were only

Table 8
Statistics for Construct Validity Models With Executive Functioning Variables

Variable	Model A	Model B	Model C	Model D
Fit statistics				
χ^2	7.52	355.71	320.67	292.73
df	5	160	158	144
CFI	.999	.987	.989	.990
RMSEA	.044	.069	.063	.063
Loadings on executive functioning (EF) construct				
EF → WCST no. errors	.54*	.54*	.53*	.10
EF → Tower (3-disk)	.34*	.39*	.42*	−.14
EF → Verbal fluency	−.62*	−.61*	−.18	−.27
EF → Figural fluency	−.69*	−.71*	−.71*	−.12
EF → Connections diff.	.50*	.46*	.46*	.20
Correlations with other constructs				
EF ←→ gF		−.94* (−.92, −.95)	−.96* (−.95, −.97)	
EF ←→ Memory		−.77* (−.72, −.82)	−.78* (−.73, −.82)	
EF ←→ Speed		−.86* (−.83, −.89)	−.85* (−.81, −.88)	
EF ←→ Vocabulary		−.43* (−.33, −.52)	−.31* (−.20, −.42)	

	gF	Memory	Speed	Vocabulary
Loadings on other constructs				
Model C				
WCST no. errors				
Tower (3-disk)				
Verbal fluency			.32*	.36*
Figural fluency				
Connections diff.				
Model D				
WCST no. errors	−.42*	−.05	−.07	−.07
Tower (3-disk)	−.38*	−.31*	.14	.20*
Verbal fluency	.17	−.07	.41*	.36*
Figural fluency	.43*	.01	.36*	−.05
Connections diff.	−.24*	−.09	−.17	−.03

Note. Values in parentheses adjacent to correlations are 95% confidence intervals created from the Fisher *r*-to-*z* transformation. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; WCST = Wisconsin Card Sorting Test; gF = fluid intelligence; diff. = difference.
* $p < .01$.

in the moderate range, indicating that from the perspective of these analyses the construct also had at least some discriminant validity.

However, the Model D analyses revealed that the relations among the variables postulated to represent inhibition were not significantly different from zero when they were considered in the context of relations to other constructs, suggesting that convergent validity was lacking. Furthermore, many of the relations between the variables and other constructs were statistically significant, suggesting that the inhibition construct had low discriminant validity. Although large in absolute magnitude, the standardized coefficient of .77 for the Stroop difference measure was not significantly different from zero because of large variability (i.e., the standard error for the unstandardized regression coefficient increased from 0.23 in Model B to 2.15 in Model D). This was likely attributable to multicollinearity because of the strong relations between the inhibition construct and the cognitive ability constructs.

Results from the analyses of the variables hypothesized to represent updating are summarized in Table 10. Although significantly greater than zero, the standardized coefficient for the loading of the digit-monitoring variable on the hypothesized updating construct in Model A was low, and consequently this variable was deleted from the subsequent analyses. The remaining three variables had significant loadings on the hypothesized updating construct in Model B. However, the correlation of this construct with the gF construct was very high, indicating little discriminant validity. The correlation between the updating and gF constructs was reduced in Model C when the variable with the highest loading (i.e., *n*-back 2) was allowed to load on other constructs, but that variable was then no longer related to the hypothesized updating construct.

In the analyses with Model D, all of the variables had statistically significant loadings on the gF construct, and none of the loadings on the hypothesized updating construct were statistically different from zero. The standardized coefficient for the keeping track variable was large in absolute magnitude but, probably

Table 9
Statistics for Construct Validity Models With Inhibition Variables

Variable	Model A	Model B	Model C	Model D
Fit statistics				
χ^2	1.97	319.31	286.36	280.84
df	3	125	124	117
CFI	1.00	.985	.987	.987
RMSEA	.000	.077	.071	.073
Loadings on inhibition (Inh) construct				
Inh → Stroop I − N diff.	.96*	.75*	.82*	.77
Inh → Read with distr. avg. *SD*	.50*	.64*	.53*	.22
Inh → Anti-cue avg. *SD*	.28*	.30*	.32*	.14
Inh → Stop signal @ 400	−.12			
Correlations with other constructs				
Inh ←→ gF		−.73* (−.67, −.78)	−.66* (−.59, −.73)	
Inh ←→ Memory		−.66* (−.59, −.73)	−.59* (−.51, −.66)	
Inh ←→ Speed		−.78* (−.73, −.82)	−.74* (−.68, −.79)	
Inh ←→ Vocabulary		−.32* (−.21, −.43)	−.15 (−.03, −.27)	

	gF	Memory	Speed	Vocabulary
Loadings on other constructs				
Model C				
Stroop I − N diff.				
Read with distr. avg. *SD*				−.34*
Anti-cue avg. *SD*				
Model D				
Stroop I − N diff.	−.26*	−.13	−.36*	.09
Read with distr. avg. *SD*	−.10	−.07	−.31*	−.30*
Anti-cue avg. *SD*	−.20	.18	−.24*	.02

Note. Values in parentheses adjacent to correlations are 95% confidence intervals created from the Fisher *r*-to-*z* transformation. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; I = incongruent; N = neutral; diff. = difference; distr. = distraction; gF = fluid intelligence; avg. = average.
* $p < .01$.

because of multicollinearity, so was the variability, and consequently it was not significantly different from zero. The implication from the Model D results is that these variables primarily reflect the gF construct and have very little residual variance in common.

Table 11 contains results from the analyses of the variables hypothesized to represent time-sharing aspects of executive control. All of the variables had statistically significant loadings on the hypothesized time-sharing construct in Models A, B, and C, suggestive of convergent validity. However, the correlations of that construct with the gF construct in Models B and C were substantially larger than the loadings of the variables on the time-sharing construct, implying that the construct had weak discriminant validity.

In the analyses with Model D, none of the variables had a significant loading on the hypothesized time-sharing construct, but they each had one or more significant loadings on the constructs representing other cognitive abilities. This pattern suggests that there is little evidence for a distinct time-sharing construct because the variables appear to reflect different combinations of other constructs and have little residual variance in common.

Two additional sets of analyses were conducted to explore possible interpretations of the weak evidence for construct validity in the results summarized in Tables 8 through 11. For example, it is conceivable that the relations among the constructs were artificially inflated owing to the relation of age to most of the variables. This possibility was examined by partialing the linear and quadratic effects of age from all variables and then repeating the analyses on the age-partialed residuals. Results from these analyses were very similar to those with the original data, as the variables had the same pattern of loadings on the constructs, and the correlations among the constructs closely resembled those in Tables 8 through 11. In particular, the correlations between the cognitive ability constructs ranged from .46 to .64, but the correlations with the gF construct were .89 for the neuropsychological executive functioning construct, .89 for the updating construct, and .81 for the time-sharing construct. It therefore does not appear that the results reported above are attributable to a spurious age-induced covariation among the variables or constructs.

Another possibility in light of the negative results with all of the Model D analyses is that this model is too stringent, such that it is unlikely that it would yield evidence of construct validity with any

Table 10
Statistics for Construct Validity Models With Updating Variables

Variable	Model A	Model B	Model C	Model D
Fit statistics				
χ^2	5.34	305.49	295.79	293.41
df	2	125	123	117
CFI	.999	.987	.988	.988
RMSEA	.080	.075	.074	.076
Loadings on updating (Upd) construct				
Upd → Keeping track	.53*	.43*	.50*	−.92
Upd → Matrix monitoring	.72*	.64*	.78*	−.17
Upd → *n*-back 2	−.53*	−.67*	.00	.01
Upd → Digit monitoring	.18*			
Correlations with other constructs				
Upd ←→ gF		.93* (.91, .94)	.76* (.70, .81)	
Upd ←→ Memory		.72* (.66, .77)	.59* (.51, .66)	
Upd ←→ Speed		.79* (.74, .83)	.63* (.55, .70)	
Upd ←→ Vocabulary		.30* (.19, .41)	.27* (.16, .38)	
	gF	Memory	Speed	Vocabulary
Loadings on other constructs				
Model C				
Keeping track				
Matrix monitoring				
n-back 2	−.45*		−.25*	
Model D				
Keeping track	.29*	.11	.03	.02
Matrix monitoring	.48*	.01	.17	−.02
n-back 2	−.46*	−.06	−.22*	.06

Note. Values in parentheses adjacent to correlations are 95% confidence intervals created from the Fisher *r*-to-*z* transformation. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; gF = fluid intelligence.
* $p < .01$.

combination of variables. To test this interpretation, each of the cognitive ability constructs was separately examined with Model D. For example, the matrix reasoning, letter sets, spatial relations, paper folding, and form boards variables were postulated to be related to a gF construct, in addition to constructs representing episodic memory, perceptual speed, and vocabulary. With all four cognitive ability constructs the loadings of the variables on the target construct were significantly different from zero, and for all but one variable the absolute magnitude of the loading on the target construct was larger than the loading of the variable on any other construct. These results therefore suggest that it is not unrealistic to assess construct validity with the model represented in Panel D of Figure 2.

A final analysis was conducted to examine the dimensionality of individual differences in the executive functioning and gF variables. That is, an exploratory factor analysis was conducted on all of the variables hypothesized to assess executive functioning, inhibition, updating, time sharing, and gF. The primary result from this analysis was that the scree plot indicated a marked discontinuity between one and two factors (i.e., eigenvalues of 6.35 vs. 1.10 and smaller). This suggests that a sizable proportion of the shared variance among these variables could be accounted for by a single factor, which is consistent with the large correlations between the various constructs reported above.

Mediation of Age–Cognition Relations

Table 12 summarizes the major results of mediation analyses based on the model portrayed in Figure 3, with the three cognitive abilities having negative relations to age serving as the cognitive constructs and the four constructs hypothesized to represent aspects of executive functioning serving as potential mediators. Within each of the three sections of the table, the first row contains the standardized coefficients for the linear (age) and quadratic (age squared) age relations corresponding to the total effects of age on the cognitive ability construct.[5] The columns from left to right contain, respectively, the standardized coefficients for the linear

[5] To minimize multicollinearity of the age and age-squared terms, the age variable was first centered following the guidelines in Cohen and Cohen (1983).

Table 11
Statistics for Construct Validity Models With Time-Sharing Variables

Variable	Model A	Model B	Model C	Model D
Fit statistics				
χ^2	0	320.90	314.24	296.55
df	0	125	124	117
CFI	1.00	.984	.985	.986
RMSEA	0	.078	.077	.077
Loadings on time-sharing (TS) construct				
TS → PA & track avg. *SD*	.36*	.36*	.41*	.13
TS → *n*-back & drive avg. *SD*	.56*	.46*	.51*	.28
TS → Connect & count avg. *SD*	.35*	.43*	.14*	.18
Correlations with other constructs				
TS ←→ gF		−.87* (−.84, −.90)	−.84* (−.80, −.87)	
TS ←→ Memory		−.52* (−.43, −.60)	−.42* (−.32, −.52)	
TS ←→ Speed		−.75* (−.69, −.80)	−.59* (−.51, −.66)	
TS ←→ Vocabulary		−.36* (−.25, −.46)	−.23* (−.11, −.34)	

Variable	gF	Memory	Speed	Vocabulary
Loadings on other constructs				
Model C				
PA & track avg. *SD*				
n-back & drive avg. *SD*				
Connect & count avg. *SD*			−.31*	
Model D				
PA & track avg. *SD*	−.52*	.28*	−.06	.07
n-back & drive avg. *SD*	−.46*	.08	−.07	.05
Connect & count avg. *SD*	−.06	.12	−.40*	−.24*

Note. Values in parentheses adjacent to correlations are 95% confidence intervals created from the Fisher *r*-to-*z* transformation. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; PA = paired associates; avg. = average; gF = fluid intelligence.
* $p < .01$.

and quadratic relations of age to the mediator, for the relation of the mediator to the cognitive construct, and for the direct linear and quadratic relations of age to the cognitive construct.

The values in the last two columns are most relevant to the mediation interpretation because mediation might be inferred to be operating to the extent that the direct age–cognition relations are smaller than the total age–cognition relations. Inspection of these values reveals that all of the direct age effects were smaller than the corresponding total age effects. Furthermore, in several cases, such as when the executive functioning or updating constructs were treated as mediators of the gF construct, the direct age relations were close to zero, which indicates that there were little or no unique age-related effects on the cognitive construct.

Although the results in Table 12 are consistent with the expected mediation pattern, they cannot necessarily be interpreted as evidence that one construct is a true mediator of the age-related effects on the other construct because the high values in the middle column suggest that the two constructs may not be distinct. That is, if the constructs represent nearly the same dimension of variation, as is likely the case when the correlation between them is very high, then the analyses may in effect be partialing the construct from itself, in which case it is not meaningful to refer to mediation.

Discussion

Many neuropsychological variables hypothesized to reflect executive functioning were originally grouped together because prior studies found them to be sensitive to damage in a particular brain region, namely, the frontal lobes. Although the classification may be meaningful in that context, particularly if it is accompanied by convincing patterns of specificity as well as sensitivity, one should not automatically assume that the same grouping is necessarily applicable to a population of healthy adults. More generally, if variables do not have construct validity, in the sense that they represent what they are intended to represent, then results involving those variables may not be interpretable.

Analyses in the current study revealed that variables assumed to reflect executive functioning or aspects of executive control have moderate loadings on the hypothesized constructs, which is consistent with the existence of convergent validity. However, by itself this pattern is not very informative, because most cognitive variables are positively correlated with one another. To provide convincing evidence of construct validity it is also necessary to examine discriminant validity to establish that the construct, or the variables assumed to represent the construct, have much weaker

Table 12
Standardized Coefficients for Executive Functioning Constructs as Mediators of Age-Related Effects on Different Cognitive Abilities

Mediator	Age→mediator Linear	Age→mediator Quadratic	Mediator→cognition	Age→cognition Linear	Age→cognition Quadratic
Cognition = gF					
Total age effects				−.45*	−.20*
Executive functioning	.51*	.32*	−1.01*	.06	.12*
Inhibition	.51*	.20*	−.60*	−.14	−.08
Updating	−.49*	−.21*	.92*	.00	−.01
Time sharing	.43*	.26*	−.83*	−.09	.02
Cognition = memory					
Total age effects				−.46*	−.23*
Executive functioning	.50*	.32*	−.69*	−.12	−.01
Inhibition	.51*	.18*	−.41*	−.26*	−.16*
Updating	−.49*	−.22*	.59*	−.18*	−.11
Time sharing	.43*	.27*	−.33*	−.32*	−.15
Cognition = speed					
Total age effects				−.57*	−.17*
Executive functioning	.49*	.32*	−.79*	−.19*	.09
Inhibition	.51*	.20*	−.60*	−.26*	−.05
Updating	−.49*	−.21*	.65*	−.25*	−.04
Time sharing	.38*	.30*	−.65*	−.32*	.03

Note. gF = fluid intelligence.
* $p < .01$.

relations to other constructs. In the case of the executive functioning, updating, and time-sharing constructs, the correlations with the gF construct were very high (i.e., greater than .85), implying that nearly the same dimension of individual differences was captured with what are often assumed to be different constructs. Stated somewhat differently, a very similar rank-ordering of individuals is produced by sorting individuals on what is common to tests such as matrix reasoning, letter series, spatial relations, paper folding, and form boards, and by sorting individuals on what is common to neuropsychological variables such as measures from the WCST, Tower of Hanoi, verbal and figural fluency, and the difference between connecting items in same or alternating sequences.

It is possible that somewhat different results might have been found with a larger sample (and greater power) or with other combinations of variables. Nevertheless, the current results indicate that with a variety of commonly used variables there is considerable overlap of the individual-differences variance in constructs hypothesized to reflect executive functioning and aspects of executive control with the individual-differences variance in gF. The discovery of weak to nonexistent evidence for construct validity with several frequently used neuropsychological and cognitive process variables raises concerns that researchers who assume that they are examining executive functioning or executive control may not actually be studying the hypothesized constructs. That is, instead of investigating aspects specific to executive functioning or executive control, the reliable variance in the target variables may represent combinations of other cognitive constructs, such as gF, episodic memory, perceptual speed, or vocabulary.

There have been many prior suggestions of a relation between executive functioning and the general factor in intelligence (e.g., Burgess, 1997; Denckla, 1995; Duncan, 1994; Duncan et al., 1995, 1996, 1997, 2000; Mittenberg et al., 1989; Obonsawin et al., 2002; Rabbitt, 1997; Schretlen et al., 2000). However, the current study provides empirical evidence that, at least in a sample of healthy normal adults, the constructs are nearly identical with respect to the pattern of individual differences. Although they interpreted the result somewhat differently, Miyake et al. (2001) also found a large correlation (i.e., .90) between a construct based on two of the variables used in this study to assess gF and an executive functioning construct based on a Tower of Hanoi variable and a variable from a random generation task. The updating and time-sharing constructs were also found to be very closely related to the gF construct in the current study. The relation with updating is consistent with several reports in the psychometric literature (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen & Christal, 1990; Miyake et al., 2001; Suss, Oberauer, Wittmann, Wilhelm, & Schulze, 2002),[6] and the time-sharing relation supports speculations about a close connection between time sharing and Spearman's *g* (e.g., Ben-Shakhar & Sheffer, 2001; Duncan, Williams,

[6] Ackerman, Beier, and Boyle (2002) recently questioned the association between working memory and gF and suggested that perceptual speed may also be involved in this relation. However, the correlations in Table 10 are only partially consistent with the Ackerman et al. (2002) interpretation because the updating (working memory) construct was more closely related to the gF construct than it was to the perceptual speed construct.

Nimmo-Smith, & Brown, 1992; Hunt & Lansman, 1982; Stankov, 1988). However, by including multiple variables and constructs in the same study, the current project provides a broader perspective on the interrelations among variables and constructs than that available from earlier studies.

One of the major goals of this project was to investigate the potential mediational role of executive functioning in age–cognition relations. Although outcomes like those summarized in Table 12 have been interpreted as evidence in support of mediational hypotheses, there are at least three reasons to be cautious about this type of inference. First, a pattern of attenuated age–cognition relations is only consistent with the mediation hypothesis; it should not be considered confirmatory. That is, the analyses primarily evaluate statistical independence, and although lack of independence is a necessary condition for mediation, it is not sufficient, because there are many reasons why variables or constructs might not be independent of one another.[7]

Second, it is not meaningful to refer to mediation if the explanatory and the to-be-explained constructs have substantial overlap with respect to the dimension of individual differences that they represent. In the case of the executive functioning and gF constructs, the overlap is nearly complete, and thus rather than the results implying that one construct mediates age-related effects on the other, they may simply reflect the consequences of partialing a construct from itself.

And third, for mediation to be plausible one must assume that the hypothesized mediator (e.g., some aspect of executive functioning) is more primitive or fundamental with respect to age-related influences than the criterion construct (e.g., gF). Although theoretical arguments can be provided to justify why particular variables or constructs should be considered to have a privileged status, causal priority is difficult to determine when the variables are all based on behavioral observations at the same point in time. Unfortunately, the most convincing type of research for this purpose, involving random assignment to intervention and nonintervention groups with long-term follow-up to monitor the sequence of age-related change in all constructs, is not yet practical. Until other types of evidence are available, therefore, there may always be some ambiguity with respect to the causal sequence of different cognitive constructs.

Results from the Model D analyses can be interpreted from at least two perspectives. One perspective might view cognitive abilities as primary, in which case the results in Tables 8 through 11 would be interpreted as indicating which abilities contribute to the variation in each variable. For example, the results in Table 8 would imply that the Tower of Hanoi variable is influenced by gF, episodic memory, and vocabulary abilities; the verbal fluency variable is influenced by speed and vocabulary abilities; and the figural fluency variable is influenced by gF and speed abilities. Furthermore, because many of the variables have significant associations with two or more cognitive abilities, this perspective would suggest that few of the variables are "ability pure."

An alternative perspective might consider particular variables as representing fundamental processes that are important in the individual differences in established cognitive abilities. Because of its breadth and hypothesized centrality, there has been considerable interest in attempting to identify processes or variables that represent the "essence" of gF. The results from the Model D analyses summarized in Tables 8 through 11 provide somewhat mixed evidence for this type of "critical variable" perspective. That is, whereas some of the relations with the gF ability are consistent with earlier speculations (e.g., the updating variables all had significant relations with gF), others were not (e.g., the WCST, figural fluency, and time-sharing variables were also significantly related to gF). Furthermore, the fact that 10 of the 14 variables in Tables 8 through 11 had significant relations with gF, with 7 of them having absolute correlations greater than .35, is not consistent with the existence of a single critical process. In addition, because only a few of the variables were exclusively related to gF, explanations would also be needed for why the variables contributed to these other abilities as well.

In summary, the results of this study have both methodological and theoretical implications. One methodological implication is that researchers should not merely assume that variables reflect a particular hypothesized construct without relevant empirical evidence. Because few cognitive or neuropsychological constructs have an external criterion or "gold standard" against which they can be validated, validity will often have to be established by relying on a pattern of convergent and discriminant validity with other variables. This is not the only possible method of investigating construct validity, but it is important that some method other than subjective judgments of face validity be used to ensure that the variables under investigation actually represent the hypothesized constructs.

A second methodological implication of the current results is that to ensure cumulative progress, researchers should examine, and interpret, new variables and proposed constructs in the context of what is already known. In the current study most of the variables hypothesized to represent aspects of executive functioning were found to have moderate to strong relations with well-established cognitive ability constructs, and it is not necessarily progress to discover that a variable is merely another indicator of one or more already recognized constructs. This is not to say that relations between new variables and existing constructs could not be informative about the precise nature, or meaning, of the established constructs. Rather, the point is that one needs to be cautious in postulating the existence of a new construct without relevant empirical evidence that it is distinct from constructs that have already been identified. This will likely require inclusion of variables to assess established theoretical constructs when investigating variables hypothesized to represent novel processes or constructs. However, these additional efforts may not only be informative about the relations between new and old constructs but also help elucidate the basis for individual differences in what might be considered to be established constructs.

The major theoretical implication of the results of this study is that because the concept of executive functioning overlaps with gF, updating (or working memory), and time-sharing (or attentional capacity) constructs, explanations of the individual differences in any of these constructs will likely have to be broader than has been assumed to be the case in the past. Each of these constructs has been speculated to involve the frontal lobes, but an assertion of frontal lobe involvement does not by itself indicate why, or how, measures of performance on such a wide variety of seemingly different cognitive tasks tend to be correlated with one

[7] A number of methodological issues have also been raised with respect to multivariate analyses of the relations between aging and cognition, several of which are discussed in Salthouse and Ferrer-Caja (2003).

another. Several researchers have hypothesized that at least some of the covariation among variables hypothesized to represent these constructs may be associated with the efficiency of specific cognitive processes, such as goal maintenance (e.g., Duncan et al., 2000) or preservation of internal representations in the presence of distraction (e.g., Kane & Engle, 2002). However, there is currently little evidence establishing that the same critical processes are involved in different cognitive tasks or that these processes are not merely different manifestations of a more fundamental phenomenon.

One intriguing possibility is that at least some of the variance shared among the variables in this study reflects variations in the integrity of circuits responsible for communication within and across neuroanatomical regions (e.g., Cabeza, 2001; Esposito, Kirkby, Van Horn, Ellmore, & Berman, 1999; Garlick, 2002; Greenwood, 2000; O'Sullivan et al., 2001; Peters et al., 1996; Pugh & Lipsitz, 2002; Valenzuela et al., 2000). Because complications in communication could affect the integration and coordination of many different types of processing, this neural connectivity hypothesis has the potential to account for effects on many different types of cognitive variables. Furthermore, there are a number of plausible sources of individual differences, and age-related differences, in the efficiency of neural communication, such as variations in the number or density of neurons, in the quantity or balance of different neurotransmitters, in the density of synapses, or in the degree of myelination.

Finally, what do the results of this study imply about the constructs of executive functioning and inhibition, updating, and time-sharing aspects of executive control? We begin by stating what the results do not mean. First, they do not mean that these constructs have been explained because they have been linked to gF, because that construct is still not well understood and may not be more fundamental with respect to causal sequence or neurobiological substrate. And second, these results do not invalidate relations that have been established involving variables hypothesized to represent executive functioning or aspects of executive control, such as those with damage in particular neuroanatomical regions. Rather, the results of this study suggest that researchers interested in investigating these and other cognitive or neuropsychological constructs should recognize that they are probably studying only one aspect of a larger phenomenon and that it may be misleading to assume that they have isolated something novel and distinct without first obtaining relevant empirical evidence.

As with all studies, the current study has a number of limitations. For example, the sample consisted of normal healthy adults who were functioning at fairly high levels, and different correlational patterns might be found with adults with brain damage, or with other special populations. Furthermore, although a large number of variables were examined, the list was not exhaustive and the patterns of covariation could be different with other combinations of variables and constructs. It is also possible that alternative analytical methods might yield somewhat different outcomes. Finally, the procedures may not be applicable to all populations, because the analytical procedure requires reliable estimates of many variables from a moderately large sample of individuals. To illustrate, in the current study each of 261 adults participated for approximately 6 hr of individual testing, for a total of over 1,560 hr, and this amount of testing may not be practical in many situations.

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.

Albert, M. S., & Kaplan, E. (1980). Organic implications of neuropsychological deficits in the elderly. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging: Proceedings of the George A. Talland Memorial Conference* (pp. 403–432). Hillsdale, NJ: Erlbaum.

Allamanno, N., Della Sala, S., Laiacona, M., Pasetti, C., & Spinnler, H. (1987). Problem solving ability in aging and dementia: Normative data on a non-verbal test. *International Journal of Neurological Science, 8,* 111–120.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13,* 909–922.

Andres, P., & van der Linden, M. (2001). Age-related differences in supervisory attentional system functions. *Journal of Gerontology: Psychological Science, 55B,* P373–P380.

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 user's guide.* Chicago: SmallWaters Corporation.

Arbuckle, T. Y., & Gold, D. P. (1993). Aging, inhibition, and verbosity. *Journal of Gerontology: Psychological Sciences, 48,* P225–P232.

Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology, 5,* 307–320.

Axelrod, B. N., & Henry, R. R. (1992). Age-related performance on the Wisconsin Card Sorting, Similarities, and Controlled Oral Word Association Tests. *The Clinical Neuropsychologist, 6,* 16–26.

Baddeley, A. D. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society, 4,* 523–526.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12,* 12–21.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1997). *Differential Aptitude Test.* San Antonio, TX: Psychological Corporation.

Ben-Shakhar, G., & Sheffer, L. (2001). The relationship between the ability to divide attention and standard measures of general cognitive abilities. *Intelligence, 29,* 293–306.

Birren, J. E. (1955). Age changes in speed of responses and perception and their significance for complex behavior. In *Old age in the modern world* (pp. 235–247). Edinburgh, Scotland: Livingstone.

Bisiacchi, P. S., Sgaramella, T. M., & Farinello, C. (1998). Planning strategies and control mechanisms: Evidence from closed head injury and aging. *Brain and Cognition, 37,* 113–116.

Bolla, K. L., Lindgren, K. N., Bonaccorsy, C., & Bleecker, M. L. (1990). Predictors of verbal fluency (FAS) in the healthy elderly. *Journal of Clinical Psychology, 46,* 623–628.

Boone, K. B., Miller, B. L., & Lesser, I. M. (1993). Frontal lobe cognitive functions in aging: Methodologic considerations. *Dementia, 4,* 232–236.

Boone, K. B., Panton, M. O., Gorsuch, R. L., Gonzalez, J. J., & Miller, B. L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Archives of Clinical Neuropsychology, 13,* 585–595.

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 and Motor Skills, 84,* 1187–1197.

Bryan, J., Luszcz, M. A., & Pointer, S. (1999). Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies. *Aging, Neuropsychology, and Cognition, 6,* 273–287.

Bunce, D. J., Warr, P. B., & Cochrane, T. (1993). Blocks in choice responding as a function of age and physical fitness. *Psychology and Aging, 8,* 26–33.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, England: Psychology Press.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4,* 547–558.

Cabeza, R. (2001). Functional neuroimaging of cognitive aging. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 331–377). Cambridge, MA: MIT Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Charness, N. (1987). Component processes in bridge bidding and novel problem-solving tasks. *Canadian Journal of Psychology, 41,* 223–243.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging, 6,* 533–541.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183.

Crawford, J. R., Bryan, J., Luszcz, M. A., Obonsawin, M. C., & Stewart, L. (2000). The executive decline hypothesis of cognitive aging: Do executive deficits qualify as differential deficits and do they mediate age-related memory decline? *Aging, Neuropsychology, and Cognition, 7,* 9–31.

Daigneault, S., Braun, C. M. J., & Whitaker, H. A. (1992). Early effects of normal aging in perseverative and non-perseverative prefrontal measures. *Developmental Neuropsychology, 8,* 99–114.

Davis, H. P., Cohen, A., Gandy, M., Columbo, P., vanDusseldorp, G., Simolke, N., & Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience, 104,* 288–297.

de Jong, R. (2001). Adult age differences in goal activation and goal maintenance. *European Journal of Cognitive Psychology, 13,* 71–89.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan executive function system.* New York: Psychological Corporation.

Delis, D. C., Kramer, J., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test—Second Edition.* San Antonio, TX: Psychological Corporation.

Della Sala, S., Gray, C., Spinnler, H., & Trivelli, C. (1998). Frontal lobe functioning in man: The riddle revisited. *Archives of Clinical Neuropsychology, 13,* 663–682.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism. Toward a unified theory of cognitive development and aging. *Developmental Review, 12,* 45–75.

Denckla, M. B. (1995). A theory and model of executive function: A neuropsychological perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263–278). Baltimore: Brookes Publishing.

Duncan, J. (1994). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33,* 261–268.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257–303.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobes after head injury: Unity and diversity of function. *Cognitive Neuropsychology, 14,* 713–741.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000). A neural basis for general intelligence. *Science, 289,* 457–460.

Duncan, J., Williams, P., Nimmo-Smith, I., & Brown, I. (1992). The control of skilled behavior: Learning, intelligence and distraction. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 322–341). Cambridge, MA: MIT Press.

Earles, J. L., Connor, L. T., Frieske, D., Park, D. C., Smith, A. D., & Zwahr, M. (1997). Age differences in inhibition: Possible causes and consequences. *Aging, Neuropsychology, and Cognition, 4,* 45–57.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

E-Prime. (2002). [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31,* 17–28.

Esposito, G., Kirkby, B. S., Van Horn, J. D., Ellmore, T. M., & Berman, K. F. (1999). Context-dependent, neural system-specific neurophysiological concomitants of ageing: Mapping PET correlates during cognitive activation. *Brain, 122,* 963–979.

Ferrer-Caja, E., Crawford, J. R., & Bryan, J. (2002). A structural modeling examination of the executive decline hypothesis of cognitive aging through reanalysis of Crawford et al.'s (2000) data. *Aging, Neuropsychology, and Cognition, 9,* 231–249.

Foster, J. K., Black, S. E., Buck, B. H., & Bronskill, M. J. (1997). Ageing and executive functions: A neuroimaging perspective. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 117–134). East Sussex, England: Psychology Press.

Fristoe, N. M., Salthouse, T. A., & Woodard, J. L. (1997). Examination of age-related deficits on the Wisconsin Card Sorting Test. *Neuropsychology, 11,* 428–436.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review, 109,* 116–136.

Gilhooly, K. J., Phillips, L. H., Wynn, V., Logie, R. H., & Della Sala, S. (1999). Planning processes and age in the five-disc Tower of London task. *Thinking and Reasoning, 5,* 339–361.

Greenwood, P. M. (2000). The frontal aging hypothesis evaluated. *Journal of the International Neuropsychological Society, 6,* 705–726.

Haaland, K. Y., Vranes, L. F., Goodwin, J. S., & Garry, P. J. (1987). Wisconsin Card Sorting Test performance in a healthy elderly population. *Journal of Gerontology, 42,* 345–346.

Hanes, K. R., Andrewes, D. G., Smith, D. J., & Pantelis, C. (1996). A brief assessment of executive control dysfunction: Discriminant validity and homogeneity of planning, set shift, and fluency measures. *Archives of Clinical Neuropsychology, 11,* 185–191.

Head, D., Raz, N., Gunning-Dixon, F., Williamson, A., & Acker, J. D. (2002). Age-related differences in the course of cognitive skill acquisition: The role of regional cortical shrinkage and cognitive resources. *Psychology and Aging, 17,* 72–84.

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.

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97–110.

Hunt, E., & Lansman, M. (1982). Individual differences in attention. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 207–254). Hillsdale, NJ: Erlbaum.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A.

(1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 362–376.

Isingrini, M., & Vazou, F. (1997). Relation between fluid intelligence and frontal lobe functioning in older adults. *International Journal of Aging and Human Development, 45,* 99–109.

Kafer, K. L., & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society, 3,* 109–119.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

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

Keys, B. A., & White, D. A. (2000). Exploring the relationship between age, executive abilities, and psychomotor speed. *Journal of the International Neuropsychological Society, 6,* 76–82.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Cognitive functions in the prefrontal cortex: Working memory and executive control. *Current Directions in Psychological Science, 6,* 185–192.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Kopelman, M. D. (1991). Frontal dysfunction and memory deficits in the alcoholic Korsakoff syndrome and Alzheimer-type dementia. *Brain, 114,* 117–137.

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9,* 491–512.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14,* 389–433.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology, 49A,* 29–50.

Levine, B., Stuss, D. T., & Milberg, W. P. (1995). Concept generation: Validation of a test of executive functioning in a normal aging population. *Journal of Clinical and Experimental Neuropsychology, 17,* 740–758.

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

Liben, D. J., Glosser, G., Malamut, B. L., Kaplan, E., Goldberg, E., Swenson, R., & Sands, L. P. (1994). Age, executive functions, and visuospatial functioning in healthy older adults. *Neuropsychology, 8,* 38–43.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 189–239). San Diego, CA: Academic Press.

Lowe, C., & Rabbitt, P. (1997). Cognitive models of ageing and frontal lobe deficits. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 39–60). Hove, England: Psychology Press.

Maki, P. M., Zonderman, A. B., & Weingartner, H. (1999). Age differences in implicit memory: Fragmented object identification and category exemplar generation. *Psychology and Aging, 14,* 284–294.

Malloy, P. F., Cohen, R. A., & Jenkins, M. A. (1998). Frontal lobe function and dysfunction. In P. J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (pp. 573–590). Washington, DC: American Psychological Association.

May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review, 6,* 142–147.

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 363–379.

McCrae, R. R., Arenberg, D., & Costa, P. T. (1987). Declines in divergent thinking with age: Cross-sectional, longitudinal and cross-sequential analyses. *Psychology and Aging, 2,* 130–137.

Mejia, S., Pineda, D., Alvarez, L. M., & Ardila, A. (1998). Individual differences in memory and executive function abilities during normal aging. *International Journal of Neuroscience, 95,* 271–284.

Mittenberg, W., Seidenberg, M., O'Leary, D. S., & DiGiulio, D. V. (1989). Changes in cerebral functioning associated with normal aging. *Journal of Clinical and Experimental Neuropsychology, 11,* 918–932.

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.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Miyake, A., & Shah, P. (2000). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Moscovitch, M., & Winocur, G. (1995). Frontal lobes, memory, and aging. *Annals of the New York Academy of Sciences, 769,* 119–150.

Mountain, M. A., & Snow, W. G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology: A review. *The Clinical Neuropsychologist, 7,* 108–118.

Neth, H. (2002). MultiTowers [Computer program]. Retrieved April 14, 2002, from www.nethde/Research/MultiTowers/MultiTowers/_E.html

Obonsawin, M. C., Crawford, J. R., Page, J., Chalmers, P., Cochrane, R., & Low, G. (2002). Performance on tests of frontal lobe function reflect general intellectual ability. *Neuropsychologia, 40,* 970–977.

O'Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. C. R., & Markus, H. S. (2001). Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline. *Neurology, 57,* 632–638.

Parkin, A. J. (1997). Normal age-related memory loss and its relation to frontal lobe dysfunction. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 177–190). Hove, England: Psychology Press.

Parkin, A. J., & Java, R. I. (1999). Deterioration of frontal lobe function in normal aging: Influences of fluid intelligence versus perceptual speed. *Neuropsychology, 13,* 539–545.

Parkin, A. J., & Java, R. I. (2000). Determinants of age-related memory loss. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 188–203). Oxford, England: Oxford University Press.

Parkin, A. J., & Lawrence, A. (1994). A dissociation in the relation between memory tasks and frontal lobe tests in the normal elderly. *Neuropsychologia, 32,* 1523–1532.

Parkin, A. J., & Walter, B. M. (1991). Aging, short-term memory, and frontal dysfunction. *Psychobiology, 19,* 175–179.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

Parkin, A. J., Walter, B. M., & Hunkin, N. M. (1995). Relationships between normal aging, frontal lobe functioning, and memory for temporal and spatial information. *Neuropsychology, 9,* 304–312.

Perfect, T. (1997). Memory aging as frontal lobe dysfunction. In M. A. Conway (Ed.), *Cognitive models of memory* (pp. 315–339). Cambridge, MA: MIT Press.

Peters, A., Rosene, D. I., Moss, M. B., Kemper, T. L., Abraham, C. R., Tigges, J., & Albert, M. S. (1996). Neurobiological bases of age-related cognitive decline in the Rhesus monkey. *Journal of Neuropathology and Experimental Neurology, 55,* 861–874.

Phillips, L. H. (1997). Do "frontal tests" measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 191–213). Hove, England: Psychology Press.

Phillips, L. H. (1999). Age and individual differences in letter fluency. *Developmental Neuropsychology, 15,* 249–267.

Phillips, L. H., & Della Sala, S. (1998). Aging, intelligence, and anatomical segregation in the frontal lobes. *Learning and Individual Differences, 10,* 217–243.

Phillips, L. H., Wynn, V., Gilhooly, S., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory, 7,* 209–231.

Pugh, K. G., & Lipsitz, L. A. (2002). The microvascular frontal-subcortical syndrome of aging. *Neurobiology of Aging, 23,* 421–431.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–38). Hove, England: Psychology Press.

Rasmussen, D. X., Zonderman, A. B., Kawas, C., & Resnick, S. M. (1998). Effects of age and dementia on the Trail Making Test. *The Clinical Neuropsychologist, 12,* 169–178.

Raven, J. (1962). *Advanced progressive matrices, Set II.* London: H. K. Lewis.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 1–90). Mahwah, NJ: Erlbaum.

Raz, N., Gunning-Dixon, F. M., Head, D., Dupuis, J. H., & Acker, J. D. (1998). Neuroanatomical correlates of cognitive aging: Evidence from structural magnetic resonance imaging. *Neuropsychology, 12,* 1–20.

Reitan, R. M., & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4,* 161–198.

Reitan, R. M., & Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *The Clinical Neuropsychologist, 9,* 50–56.

Rende, B., Rambsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology, 16,* 309–321.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, L. D., McInnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society, 4,* 474–490.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1997). A neural systems approach to the cognitive psychology of ageing using the CANTAB battery. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 215–238). Hove, England: Psychology Press.

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

Rubin, D. C. (1999). Frontal-striatal circuits in cognitive aging: Evidence for caudate involvement. *Aging, Neuropsychology, and Cognition, 6,* 241–259.

Ruff, R. M. (1996). *Ruff Figural Fluency Test.* Odessa, FL: Psychological Assessment Resources.

Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences, 48,* P29–P36.

Salthouse, T. A. (1998). Independence of age-related influences on cognitive abilities across the life span. *Developmental Psychology, 34,* 851–864.

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

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6,* 118–127.

Salthouse, T. A., & Czaja, S. J. (2000). Structural constraints on process explanations in cognitive aging. *Psychology and Aging, 15,* 44–55.

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging, 18,* 91–110.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology, 10,* 272–285.

Salthouse, T. A., Hambrick, D. Z., & McGuthry, K. E. (1998). Shared age-related influences on cognitive and noncognitive variables. *Psychology and Aging, 13,* 486–500.

Salthouse, T. A., Hancock, H. E., Meinz, E. J., & Hambrick, D. Z. (1996). Interrelations of age, visual acuity, and cognitive functioning. *Journal of Gerontology: Psychological Sciences, 51B,* P317–P330.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences, 50B,* P297–P306.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition, 30,* 572–582.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. (2000). Effects of aging on the efficiency of task switching in a variant of the Trail Making Test. *Neuropsychology, 14,* 102–111.

Salthouse, T. A., Toth, J. P., Hancock, H. E., & Woodard, J. L. (1997). Controlled and automatic forms of memory and attention: Process purity and the uniqueness of age-related influences. *Journal of Gerontology: Psychological Sciences, 52B,* P216–P228.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologica, 40,* 605–619.

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., & Barta, P. (2000). Elucidating the contribution of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. *Journal of the International Neuropsychological Society, 6,* 52–61.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in the brain. *Brain, 114,* 727–741.

Shimamura, A. P. (1984). Neuropsychological perspectives on memory and cognitive decline in normal human aging. *Seminars in the Neurosciences, 6,* 387–394.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology, 14,* 299–309.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests* (2nd ed.). New York: Oxford University Press.

Stankov, L. (1988). Single tests, competing tasks, and their relationship to the broad facets of intelligence. *Personality and Individual Differences, 9,* 25–33.

Stoltzfus, E. R., Hasher, L., Zacks, R. T., Ulivi, M. S., & Goldstein, D. (1993). Investigations of inhibition and interference in younger and older adults. *Journal of Gerontology: Psychological Sciences, 48,* P179–P188.

Suss, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261–288.

Tomer, R., & Levin, B. E. (1993). Differential effects of aging on two verbal fluency tasks. *Perceptual & Motor Skills, 76,* 465–466.

Tranel, D., Anderson, S. W., & Benton, A. (1994). Development of the concept of "executive function" and its relationship to the frontal lobes. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 125–148). Amsterdam: Elsevier.

Troyer, A. K., Graves, R. E., & Cullum, C. M. (1994). Executive functioning as a mediator of the relationship between age and episodic memory in healthy aging. *Aging and Cognition, 1,* 45–53.

Vakil, E., & Agmon-Ashinazi, D. (1997). Baseline performance and learning rate of procedural and declarative memory tasks: Younger versus older adults. *Journal of Gerontology: Psychological Sciences, 52,* 229–234.

Valenzuela, M. J., Sachdev, P. S., Wen, W., Shnier, R., Brodaty, H., & Gillies, D. (2000). Dual voxel proton magnetic resonance spectroscopy in the healthy elderly: Subcortical-frontal axonal n-acetylaspartate levels are correlated with fluid cognitive abilities independent of structural brain changes. *Neuroimage, 12,* 747–756.

Veroff, A. E. (1980). The neuropsychology of aging: Qualitative analysis of visual reproduction. *Psychological Research, 41,* 259–268.

Ward, G., Roberts, M. J., & Phillips, L. H. (2001). Task-switching costs, Stroop-costs, and executive control: A correlational study. *Quarterly Journal of Experimental Psychology, 54A,* 491–511.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Whelihan, W. M., & Lesher, E. L. (1985). Neuropsychological changes in frontal function with aging. *Developmental Neuropsychology, 1,* 371–380.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock–Johnson Psycho-Educational Battery—Revised.* Allen, TX: DLM Teaching Resources.

Yntema, D. B., & Mueser, G. E. (1960). Remembering the present states of a number of variables. *Journal of Experimental Psychology, 60,* 18–22.

Received November 6, 2002
Revision received March 18, 2003
Accepted May 26, 2003 ■