This is to certify that the

thesis entitled

UNITY, RELIABILITY, AND STRUCTURAL STABILITY OF
EXECUTIVE FUNCTION IN A NORMAL AGING POPULATION

presented by

Mark Lawrence Ettenhofer

has been accepted towards fulfillment
of the requirements for

M.A. degree in Psychology

Major professor

Date 5/2/03

UNITY, RELIABILITY, AND STRUCTURAL STABILITY OF EXECUTIVE FUNCTION IN A NORMAL AGING POPULATION

By

Mark Lawrence Ettenhofer

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

2003

ABSTRACT

UNITY, RELIABILITY AND STRUCTURAL STABILITY OF
EXECUTIVE FUNCTION IN A NORMAL AGING POPULATION

By

Mark Lawrence Ettenhofer

Available evidence suggests that the decline of executive functions is a common part of the normal aging process (Garden, Philips, & MacPherson, 2001; Wecker, et al., 2000; Schretlen, et al., 2000; Bryan & Luszcz, 2000; Robbins, et al., 1998; West, 1996). However, many questions have been raised about the reliability, structure, and stability of executive functions. This study addressed these questions via the collection of neuropsychological data at two time points from a sample of 151 individuals aged 55 and older. This data included five common measures of executive function (Trail Making Test, Stroop Color-Word Test, Wisconsin Card Sorting Test, Letter Fluency, and Category Fluency). Exploratory factor analyses found that a single factor accounted for greater than 50% of the variance in four measures of executive function at both time points; Stroop Color-Word Test was dropped from analysis. Structural analyses of the longitudinal model found that both the regression weights of the executive measures and the variance of the error terms could be held constant across both time points without a significant reduction in model fit, indicating a large degree of longitudinal stability of executive function. Structural analyses of this longitudinal model found that the effect of age on executive function increased significantly from Time 1 to Time 2 ($\chi^2(1, \underline{n} = 151) = 9.95, \underline{p} < .01$).

ACKNOWLEDGEMENTS

This research was supported by a Student Award Program Grant from the Blue Cross Blue Shield of Michigan Foundation, and a grant from the Pearl Aldrich Endowment in Gerontology. I also owe many thanks to Norman Abeles for serving as my mentor and supervising the collection of the data for this study long before my arrival in East Lansing, and to Zach Hambrick for the countless hours of assistance he provided in the preparation of this manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Popular beliefs about the effects of aging vary across cultures. Whereas Chinese individuals generally ascribe wisdom, power, and prestige to advancing age, those living in the United States often believe that aging inevitably leads to senility and powerlessness (Silverman, Hecht, & McMillin, 2000). Scientific studies have demonstrated that although some individuals seem to progress into old age without developing any significant mental deficits (Seeman, Lusignolo, Albert, & Berkman, 2001; Silver & Perls, 2001), others may lose aspects of their memory, their visuospatial skills, or their mental speed (Ylikoski et al., 1999; Mitrushina, Uchiyama, & Satz, 1995). Additionally, the process of aging often causes subtle but important changes to an individual's "executive" abilities, including volition, inhibition, and attention (Wecker, Kramer, Wisniewski, Delis, & Kaplan, 2000).

Research examining the causes of cognitive aging has consistently demonstrated that measures of psychomotor speed can account for a large amount of detectable age-related variance in test scores, supporting the hypothesis that generalized slowing is at least partially responsible for decrements in cognitive performance with advanced age (Schretlen et al., 2000; Salthouse, Fristoe & Rhee, 1996; Salthouse, 1991). However, other researchers have proposed that the frontal lobes are disproportionately vulnerable to the effects of aging (Celine, Michel, & Espagnet, 2000; Rympa, Prabhakaren, Desmond, & Gabrieli, 2001; Goldberg, 2001), and that this deterioration is one causative factor of the "normal" decline seen in a wide variety of cognitive abilities (West, 1996; Band, Ridderinkhof, & Segalowitz, 2002). Because executive function as a whole is largely reliant upon the frontal lobes (Lezak, 1995; Band et al., 2002; Celine, Michel, &

Espagnet, 2000), cognitive functions that are normally coordinated with the aid of proper executive function may become dysregulated once frontal deterioration has progressed to a significant degree (Goldberg, 2001).

Research has consistently demonstrated the negative relationship between age and executive function (Garden, Philips, & MacPherson, 2001; Wecker et al., 2000; Schretlen et al., 2000; Bryan & Luszcz, 2000; Robbins et al., 1998; West, 1996). Although the precise definition of "executive function" is a matter of great debate, it is generally accepted that this category includes the ability to plan and implement goal-directed actions, voluntarily initiate and terminate behaviors, and purposefully direct attention and behavior (LaRue, 1992). Speaking more generally, Lezak (1995) describes executive functions as those functions that are "necessary for appropriate, socially responsible, and effectively self-serving adult conduct" (p. 507). Additionally, these functions tend to cluster together; it is relatively rare to find an individual with a deficit in one executive function without deficits in others as well (Lezak, 1995). Because the category of functions described as "executive" is rather abstract and composed of diverse elements, tasks that assess executive function vary considerably.

<u>Reliability of Executive Function</u>

Although tasks purporting to measure executive function have been used extensively in aging research, relatively little is known about the psychometric properties of these tasks, either individually or collectively (Rabbitt, 1997). Additionally, what evidence is available has led some authors to conclude that executive measures as a whole cannot be validly administered more than once (Lowe & Rabbitt, 1998; Rabbit, 1997; Phillips, 1997). Phillips (1997) states:

> The whole idea of assessing test-retest reliability in executive function is problematic because a task cannot be novel the second time around. Even conventional parallel forms of the same test do not overcome the problem: the content may be new, but the format is not. Also, how could the spontaneous generation of, say, a particular task strategy be measured twice? (p. 208)

Assertions such as these cast serious doubt on the prospect of examining changes in executive function across multiple measurement occasions. An evaluation of these beliefs is a prerequisite to the meaningful examination of age effects on executive function in a longitudinal context. In order to address these issues in this study, test-retest correlations have been computed for each of five measures of executive function. Additionally, analyses have been conducted of the longitudinal stability of factor models of executive function, in order to provide a more complete picture of the reliability of the underlying construct.

As demonstrated by Hertzog and Shaie (1986) and replicated in this study, one means of testing for longitudinal invariance in a construct such as executive function involves the analysis of a series of structural models that include data from multiple measurement occasions. First, a theoretically-based baseline model is created, and model fit is estimated. Additional constraints, such as restricting factor loadings to be constant across measurement occasions, are imposed for each of the models that follow. Individual hypotheses regarding the invariance of factor loadings, error variances, and factor variances, can then be tested by comparing the overall fit of the successive models (Hertzog & Shaie, 1986).

Executive Function as a Unified Construct

Although widespread use of the term "executive function" itself implies a large degree of convergence between the tasks designed to measure it, it is unclear to what degree executive function may be more accurately characterized as consisting of multiple independent functions. For example, Duncan, Emslie, and Williams (1996) cite many studies employing animal models that seem to indicate anatomical modularity of executive function.

Support in humans for this type of modularity was found in an exploratory factor analysis of "dysexecutive" symptoms, which included such indicators as behavioral deficits in inhibition, intentionality, executive memory, and affect (Burgess, Alderman, Evans, Emslie, & Wilson, 1998, p. 547). Five orthogonal factors were found, collectively explaining 67.2% of the variance. The authors concluded that neuropsychological assessments should include multiple measures of executive function to properly tap various components of the "dysexecutive syndrome" (Burgess et al., 1998, p. 556).

Additionally, among a sample of head-injured patients, Duncan, Johnson, Swales, and Freer (1997) found low correlations between executive measures, implying more divergence than convergence between executive functions. Supporting these results, Lamar, Zonderman, and Resnick (2002) conducted a factor analysis of a wide range of neuropsychological tasks (both executive and non-executive) in a sample of nondemented older adults, and found no evidence that any of the factors that emerged were uniquely executive in nature. In contrast, in a factor analysis of patients with frontal lesions, Della Sala, Gray, Spinnler, and Trivelli (1998) found a single factor that accounted for 53% of the variance among executive measures, as well as an intercorrelation matrix that averaged $\underline{r} = .48$. The conclusion that these authors drew was that a unified model of executive function continues to be the best available conceptualization.

Other studies have produced more mixed results. One factor analysis including multiple indices from four measures of executive function found three separate but modestly correlated factors (Boone, Ponton, Gorsuch, Gonzalez, & Miller, 1998). These results, drawn from a sample of patients referred for neuropsychological evaluation, seem to suggest some degree of both unity and diversity in executive function. Likewise, a confirmatory factor analysis of nine executive measures in a college sample found that a model with three distinct executive factors fit the available data better than a single-factor model (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). However, much like the previous study, the authors in this study found moderate correlations between these three factors, suggesting a significant degree of communality between measures as well.

Clearly, the issue of unity versus diversity of executive functions remains unresolved. Further evidence is necessary in order to conclusively determine if the implicit one-factor model of executive function that continues to influence much clinical judgment is adequate, or if more sophisticated models of executive functioning are necessary. This study includes a series of factor analyses involving five common measures of executive function in order to provide evidence that may help resolve this discrepancy.

Evidence of Age-Related Frontal/Executive Decline

Although the individual neuropsychological profiles of older adults are quite diverse (Ylikoski et al., 1999), in aggregate form the data from neuropsychological testing paint a clear picture of executive decline with increased age (Garden, Philips, &

MacPherson, 2001; Wecker et al., 2000; Schretlen et al., 2000; Bryan & Luszcz, 2000; Robbins et al., 1998; West, 1996). As a whole, neuropsychological tests of executive function demonstrate a great deal of sensitivity to frontal lobe damage, particularly in the prefrontal regions (Lezak, 1995). To the degree that these executive tests are sensitive to frontal dysfunction, the consistent findings of age-related decline that they yield demonstrate changes in the functional capacity of the frontal lobes attributable to aging. Specific examples of executive tests that have been shown to demonstrate sensitivity to frontal lobe damage include the Trail Making Test (TMT; Crowe, 1998; Arbuthnott & Frank, 2000), the Stroop Color-Word Test (SCWT; Stuss, Floden, Alexander, Levine, & Katz, 2001; West & Bell, 1997; Vendrell et al., 1995), the Wisconsin Card Sorting Test (WCST; Wang, Kakigi, & Hoshiyama, 2001; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000; Konishi et al., 1999; Lombardi et al., 1999; Raz, Gunning-Dixon, Head, Dupuis, & Acker, 1998; Berman et al., 1995; Rezai et al., 1993), and tests of verbal fluency (Lezak, 1995). Additionally, as will be addressed in more detail below, studies employing brain imaging techniques have established direct connections between age-related decline of the frontal lobes and neuropsychological tests of executive function, including the CFL (Frith, Friston, Liddle, & Frackowiak, 1991), the Wisconsin Card Sorting Test (WCST; Raz et al., 1998) and the SCWT (Milham et al., 2002).

Structural brain imaging has also proved to be a valuable resource for information regarding the effects of aging on the brain. Overall brain volume has been reported to decline steadily from age 16 to 65 (Bilger, 1997). However, an investigation comparing young adults to individuals in their 70s has demonstrated volume reductions in the frontal cortex of 10-17%, compared to volume reductions of approximately 1% in the temporal, parietal, and occipital cortices (Haug & Eggers, 1991). A similar study focusing upon the prefrontal cortex found that this area was more affected by age (volumetrically) than any other region of interest (Raz et al., 1998). Corroborating the link between age-related frontal dysfunction and performance on neuropsychological tests of executive function, this study also found that atrophy of the prefrontal cortex mediates age-related decreases in performance on the WCST. Together, these data suggest that the frontal cortex is selectively susceptible to the deleterious effects of aging, and that tests of executive function may be sensitive to these changes.

Although structural information has proven very useful, age-related changes in the brain may also occur on a more microscopic level or may be functional in nature, and may therefore remain invisible to structural brain scans. In these cases, functional brain imaging is useful in detecting irregularities or changes in metabolic activity underlying a functional deficit (Fuster, 1997). Accordingly, this technology also provides a window into the structures associated with neuropsychological task performance in healthy individuals, and differences in brain metabolism between young and old individuals. For example, an age-related reduction of up to 27% in regional cerebral blood flow (rCBF) has been demonstrated in some cortical regions, as measured by positron emission tomography (PET) (Shaw et al., 1984).

Generally speaking, reduced blood flow to a brain area is indicative of lower regional metabolism, which results from lower brain activity (Goldberg, 2001; Fuster, 1997). However, this reduction in overall brain metabolism is not uniform across different brain regions. Both cross-sectional and longitudinal studies have demonstrated that older individuals often exhibit a resting pattern of metabolic hypofrontality, in contrast to the relative hyperfrontality evident in younger adults (Goldberg, 2001; Giaquinto, 1988; Gur, Gur, Orbist, Skolnick, & Reivich, 1987; Shaw et al., 1984). This pattern seems to vary depending upon the task in question. During visual perception tasks, for example, older individuals have been consistently found to display more frontal lobe activation than their younger counterparts (Cabeza, 2001). Although it may seem counterintuitive that older adults would display relatively more regional activation for a given task, results such as these are often explained in terms of adaptive compensatory processes. As individuals age, it may become increasingly difficult to coordinate the myriad component abilities involved in even simple tasks without the recruitment of supplementary frontal areas. Consistent with this hypothesis, older adults who engage supplementary areas of the prefrontal cortex often perform better than older adults who do not, despite the fact that younger adults tend to perform better overall (Cabeza, 2001).

Functional imaging has also been used to validate various neuropsychological tasks by isolating the areas of the brain which are uniquely activated during their performance. PET examinations of letter fluency (CFL), for example, have demonstrated that this task is associated with an increase in activity in the left dorsolateral prefrontal cortex, providing evidence that this measure is a meaningful index of frontal lobe function. Similar activation of the dorsolateral prefrontal cortex was found in an fMRI study of the SCWT (Milham et al., 2002). In addition, this study found age differences in the responsiveness of the dorsolateral prefrontal cortex to the demands of the SCWT, evidence suggestive of age-dependent frontal/executive decline.

To summarize, the relationships between age, executive function, and frontal lobe activity are fairly well documented. The presence of these relationships makes greater conceptual understanding and more accurate assessment of executive function within an aging population a high priority. In this study, the longitudinal stability of the relationship between age and executive function has been examined, as it was believed that an examination of this relationship might provide more information regarding the general stability of the construct. The hypothesis that age effects on executive function would be similar across two measurement occasions was tested in the manner demonstrated by Hertzog and Shaie (1986); that is, the overall fit of a model in which the effect of age on executive function at two measurement occasions was held constant was compared to the fit of a baseline model in which the effect of age on executive function was freely estimated.

Overview and Predictions

Available evidence suggests that the decline of executive functions is a common part of the normal aging process. However, many questions have been raised about the reliability, structure, and stability of executive functions, both in general as well as within an aging population. This study was designed to address these questions via the collection of neuropsychological data from a large sample of aging individuals at two time points. First, the reliability of various measures of executive function was estimated by comparing individuals' scores at two measurement occasions. Second, factor analyses were performed in order to determine the structure of executive function within this population. Guided by this information, analyses of the longitudinal stability of executive functions were then conducted. Further analyses were conducted in order to examine the magnitude of the effect of age on executive function, and the longitudinal stability of this effect.

We expected that reliability coefficients for the measures of executive function employed in this study would fall within the moderate-to-high range, and that these coefficients would be greater for measures that utilized directly observed scores (CFL, CAT, WCST) than for measures that rely upon composite scores (SCWT, TMT). Furthermore, we expected that the results of factor analyses would be consistent with a single-factor model of executive function, and that this factor structure would be stable across the two time points. Finally, we expected the results of structural analyses to be consistent with a large and longitudinally stable effect of age upon the latent executive function variable.

METHOD

Participants

Participants consisted of 151 home-dwelling individuals aged 54 to 92 (M = 69.80, SD = 8.47) who had been recruited through local newspaper advertisements and talks given to local community groups in the greater Lansing area. These individuals were solicited to participate in educational "Mood and Memory" seminars, and also received feedback regarding their performance subsequent to all testing. Only participants who completed both testing sessions (before and after the seminar) were included. Participants whose MMSE scores suggested dementia (24 or less) and participants who exhibited moderate to severe depression on the BDI (30 or greater) or the GDS (20 or greater) were excluded from analysis (Folstein, Folstein, & McHugh, 1975; Beck, Steer, & Garbin, 1988). These cutoffs help ensure that the sample is representative of the "normal" aging population. Gender and years of education were also monitored in order to determine sample characteristics. Detailed demographic characteristics of the sample are presented in Table 1.

Procedure

In addition to self-report demographic information, a standard battery of neuropsychological test data was collected, including the following measures: the State-Trait Anxiety Inventory, the MMSE, CFL, CAT, the California Verbal Learning Test, TMT Part A and B, WAIS-III Digit Span, WAIS-III Visual Memory Span, WAIS-III Digit-Symbol Coding, the WCST, the SCWT, the American version of the New Adult Reading Test, the Storandt Mental Control Battery, the Benton Visual Retention Test, WAIS-III Symbol Search, and the Clock Drawing Test. This battery typically takes 90 to

120 minutes to complete. These tests were administered both immediately prior to participation and subsequent to participation in the "Mood and Memory" seminars, yielding two sets of data for each participant (Time 1 and Time 2). The amount of time that passed between Time 1 and Time 2 varied, but was always between four and eight weeks.

Measures

Many questions have been raised about the interpretation of neuropsychological test data in regard to assumptions that are often made concerning the localization of brain damage when neuropsychological test deficits exist. Although any given measure may demonstrate sensitivity to damage in particular regions of the brain, it may not always be appropriate to interpret test deficits in terms of damage to that brain region, as similar deficits could be caused by damage to other brain regions as well (Salthouse, Fristoe, & Rhee, 1996). In response to this and related issues, great care has been taken to ensure that the measures selected have demonstrated validity for their intended purpose in this project.

Screening Instruments

Mini Mental Status Exam. (MMSE; Folstein, Folstein, & McHugh, 1975): The MMSE is a brief measure of overall orientation widely used in assessments of cognitive function. The test consists of 30 items, that individually also provide rough measures of memory, visuospatial function, and the ability to follow instructions. With most populations, a cutoff score of 24 is recommended in order to screen out individuals with dementia or other clinically-meaningful cognitive problems (Lezak, 1995).

Test-retest reliability for this test has been estimated at .85 to .99 after 24-hours

(Foster, Sclan, Welkowitz, Boksay, & Seeland, 1988). This test has also been shown to correlate highly with measures intelligence, memory, and executive function (Mitrushina & Satz, 1991; Axelrod, Goldman, & Henry, 1992). Additionally, studies have show that the MMSE has adequate specificity and sensitivity for detecting moderate to severe forms of dementia (Spreen & Strauss, 1998).

Beck Depression Inventory. (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961): The BDI is a self-report measure consisting of 21 statements, each addressing a depressive symptom. The participant is asked to rate the accuracy of each item based upon their experience on a scale from 0 to 3. Total score for the BDI is the sum of all 21 numbers circled. Scores range from 0 to 63, higher scores indicating higher measured levels of depression. The following cutoff scores were identified by the authors of this test: normal: 0-9; mild depression: 10-15; mild/moderate depression: 16-19; moderate/severe depression: 20-29; severe depression: 30 or greater (Spreen & Strauss, 1998). Test-retest reliability for this measure has been estimated to be above .90 (Beck, Steer, & Garbin, 1988).

Geriatric Depression Scale. (GDS; Yesavage et al., 1983): The GDS is a self-report paper-and-pencil measure consisting of 30 items for which scoring directionality of answers (yes or no) changes pseudo-randomly. Because this test was developed for use with older adults, it focuses upon items that the authors found to be most relevant to an aging population. For example, the somatic items in the GDS have been tailored in such a way in which they are thought to be appropriate for older populations (Spreen & Strauss, 1998). The participant is asked to read each statement and circle the response that is most accurate in terms of their own experience. Total score for this test ranges from 0 to 30, higher numbers indicating higher measured levels of depression. The following cutoff scores identify graded levels of depressive symptomology: normal: 0-9; mild depression: 10-19; moderate/severe depression: 20-30 (Spreen & Strauss, 1998). Correlations between the GDS and other measures of depression are high for the Beck Depression Inventory (r = .73), the MMPI Depression Scale (r = .72), and the Hamilton Depression Scale (r = .83; Yesavage et al., 1983).

Executive Function

The Stroop Color-Word Test. (SCWT; Stroop, 1935): The Stroop Color-Word test requires participants to read through one list of words and successively name two lists of colors as rapidly as possible. Each condition has a time limit of 45 seconds. In the first condition ("word"), the list is composed of five columns of 20 color words ("blue", "red", "green", or "yellow") printed in black ink. In the second condition ("color"), the participant is required to identify the color of ink (blue, red, green, or yellow) in which a list of pseudo-words (all "xxxx") is printed. This list of pseudo-words is in the same 20 x 5 format as the words in the first trial. Finally, the third condition requires the identification of color of ink in which a list of color words ("blue", "red", "green", or "yellow") is printed, again in a 20 x 5 format. This is called the "interference" condition, because there is a discrepancy between the color of ink and the color word itself, making identifying the color more effortful and difficult. Performance on the SCWT is assessed by the difference between the number of words achieved in the interference condition and the color condition (the "interference effect"; Lezak, 1995; Spreen & Strauss, 1998).

It is believed that the SCWT measures the degree to which participants are able to shift attention and inhibit prepotent responses, two executive functions associated with the frontal lobes (Spreen & Strauss, 1998; West, 1996). The validity of this assumption is supported by Vendrell, et al. (1995), who found a significant difference in number of interference condition errors committed between frontal lobe patients and matched controls. Likewise, Stuss, et al. (2001) found that among 51 patients with frontal and non-frontal brain lesions and 26 normal controls, only those patients with frontal lesions exhibited significant impairment on the SCWT. Psychophysiological evidence confirms the importance of frontal functions to performance on the SCWT as well. Pairing electroencephalography (EEG) with the SCWT, West & Bell (1997) found a relationship between the magnitude of the interference effect and age-related decline of the anterior attention system. Finally, support for use of the SCWT as a measure of prefrontal/executive function within an aging population comes from a study employing functional magnetic resonance imaging, which found both poorer performance on the SCWT and decreased responsiveness of the dorsolateral prefrontal cortex among older individuals (Milham et al., 2002). Test-retest reliabilities for the word, color, and interference conditions of the SCWT have been estimated at .90, .83, and .91, respectively (Spreen & Strauss, 1998).

Wisconsin Card Sorting Test. (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993): The WCST is a test that requires the participant to sort up to cards consecutively under one of four "key cards". On each card are printed one to four triangles, stars, crosses, or circles in red, green, yellow, or blue. The participant is not told how to sort the cards, and instead must deduce the correct principle using only the examiner's feedback. The examiner says "correct" or "incorrect" after each sort depending upon the sorting principle. This sorting principle changes each time the participant sorts 10 cards in a row correctly, but the participant is not informed of this change, and must deduce from examiner feedback what rule applies at any given time. The possible dimensions by which the cards can be sorted are the form (shape) of the symbols, the color of the symbols, and the number of symbols on the card.

This test was originally designed to measure abstract reasoning and concept formation abilities in response to a changing context (set-shifting), executive functions presumed to be reliant upon the frontal lobes (Barcelo, 2001). Since its introduction, the WCST has endured a great deal of controversy, and remains one of the most widely used neuropsychological tests in existence. As such, a great deal of research has been dedicated to validating the WCST for the purpose of measuring executive and frontal lobe function.

Initial support came from findings that individuals with prefrontal lesions tend to exhibit a greater number of perseverative errors on this test (Anderson, Damasio, Jones, & Tranel, 1991). Further support for the ability of the WCST to measure frontal lobe function has come from studies employing neuroimaging techniques, including: a SPECT imaging study that found that the WCST task is associated with activation in the left dorsolateral prefrontal cortex (Rezai et al., 1993), a functional magnetic resonance imaging (fMRI) study that demonstrated that both the working memory and set-shifting components of the WCST operate within the prefrontal cortex (Konishi, et al., 1999), a positron emission tomography (PET) study that demonstrated that metabolic deficits in the right dorsolateral prefrontal cortex predict greater numbers of perseverative errors (Lombardi et al., 1999), as well as multiple PET and magneto-encephalography (MEG) studies that demonstrated that performance of the WCST is associated with activation in frontal areas (Berman et al., 1995; Rogers, et al., 2000; Wang, Kakigi, & Hoshiyama, 2001). Additionally, an MRI study that demonstrated that shrinkage of the prefrontal cortex due to normal aging predicts greater numbers of perseverative errors on the WCST (Raz et. al, 1998) supports the present use of the WCST to measure prefrontal/executive function within an aging population. Lezak (1995) reports estimates of interscorer and intrascorer reliabilities on the WSCT to be .88 and .96, respectively (Lezak, 1995). Performance in this study was assessed by the total number of perseverative errors. To summarize, Heaton et al. (1993) define perseverative errors as incorrect responses that were correct in immediately preceding category, but for which potentially corrective feedback has already been provided.

Trail Making Test. (TMT; Reitan, 1958): The TMT is a paper-and-pencil test consisting of Part A and Part B. Part A requires the participant to draw a line, connecting numbered circles in sequential order. Part B requires the participant to draw a line, connecting circles with numbers or letters in alternating sequential-alphabetic order. Any mistakes made during the performance of this test corrected by the examiner as they are made. Score on each part of the TMT is determined by the time required to complete each trial. Because errors must be corrected before continuing, they augment this time score.

By virtue of its relatively simple design, performance on Part A of the Trail Making Test depends largely upon the participant's psychomotor speed and visual search abilities (Crowe, 1998; Arbuthnott & Frank, 2000). Part B, which requires the participant to maintain two mental sets (both numbers and letters) and alternate between them, places additional demands upon the participant's working memory and cognitive flexibility, two important elements of executive function (Crowe, 1998; Arbuthnott & Frank, 2000). This study employs a (Part B – Part A) derived score as a measure of executive function.

Letter Fluency (CFL) and Category Fluency (CAT). (Benton, Hamsher, & Sivan, 1983): CFL and CAT, referred to collectively as the "Controlled Oral Word Association Test" (COWAT), are tests of verbal fluency consisting of three trials. For CFL, the examiner instructs the participant in each trial to say as many words that begin with a particular letter of the alphabet as possible, excluding proper nouns as well as different forms of the same word (i.e., "bookworm" and "bookshelf"). For this study, the letters C, F, and L were used, based on the respectively high, moderate, and low frequencies of words in the English language that begin with these letters. For CAT, the examiner instructs the participant to say as many words that belong to a certain category as possible. In this study, the following categories were used: animals, vegetables, fruits.

The COWAT has been found to be a sensitive indicator of brain dysfunction, especially in the frontal lobes (Lezak, 1995). A study of the CFL using Positron Emission Tomography (PET) provided further validation for the CFL as a measure of frontal lobe function by demonstrating that generation of words in this manner was uniquely associated with activation in the left dorsolateral prefrontal cortex (Frith et al., 1991). Another study of rCBF has demonstrated that performance of both CFL and CAT activate similar, but not identical, brain regions, including left prefrontal regions (Gourovitch et al., 2000). In addition, previous studies have demonstrated that the CFL is sensitive to age effects, in that younger adults perform better than older adults by generating a greater number of words for each letter (Winocur, Moscovitch, & Stuss, 1996). Performance on the CFL and CAT was assessed by the total number of correct words generated across all three trials.

Analysis

The SPSS 10.0.5 (1999) statistical package was used to perform all correlational and exploratory factor analyses, and the AMOS 4.01 (1999) statistical package was used for all structural analyses. First, descriptive statistics and test-retest correlations of the executive function measures at Time 1 and Time 2 were computed.

Exploratory factor analyses were performed for the executive function variables at Time 1 and Time 2, respectively, in order to examine the factor structure of executive function in this population. Consistent with the guidelines outlined by Fabrigar, Wegener, MacCallum, and Strahan (1999), all analyses employed principal axis factoring, with Promax rotation for multiple-factor solutions. The first step in these exploratory factor analyses included the following variables at Time 1 and Time 2, respectively: WCST Perseverative Errors, TMT Composite Score, SCWT Interference score, CAT Total, and CFL Total. Initially, factors with eigenvalues of one or greater were included in the factor analysis. Further analyses were restricted to one-factor solutions in the interest of parsimony.

Structural analyses were then performed in which the factor loadings of the following variables on their respective latent "executive function" variables at Time 1 and Time 2 were freely estimated: WCST Perseverative Errors, TMT Composite Score, Category Fluency Total, and CFL Total (see Figure 1). As is common practice for longitudinal factor models of this type (Hertzog & Shaie, 1986), latent executive function variables and residuals were permitted to autocorrelate across time in this model (denoted

O_1) in order to achieve maximal model fit. A model similar to O_1 was then estimated, with the added restriction that all factor loadings be held constant (model O_2). A chi-square value was computed in order to determine the significance level of the difference in overall model fit between the models O_1 and O_2. A model similar to O_2 was then estimated, with the added restriction that the residual variance was held constant across time (model O_3). Because only the variance of EF remained freely estimated in this model, the chi-square comparison of O_3 with O_2 essentially tested the hypothesis that any observed differences between Time 1 and Time 2 could be attributed to changes in EF factor variance. Model O_4 was then estimated in which the hypothesis of complete longitudinal invariance was tested by adding the restriction that EF factor variance was held constant. In the next step, structural analyses were conducted on a model similar to model O_3 which allowed the observed variable of age to load freely on EF_1 and EF_2 (model O_5; see Figure 2). Model O_6 was then estimated by restricting the path coefficients of age on EF_1 and EF_2 to be held constant.

RESULTS

Reliability of Executive Measures

First, descriptive statistics and dependent-sample t-tests were computed for all executive measures (see Table 2). Mean number of WCST perseverative errors was significantly lower (indicating better performance) at Time 2 than at Time 1. Mean CFL total score was significantly higher (indicating better performance) at Time 2 than at Time 1. Mean TMT time score, Category Fluency total score, and Stoop interference score were not significantly different at Time 2 than at Time 1.

Next, test-retest correlations were computed for all executive variables (see Table 3). Test-retest correlations for CFL total score and Category Fluency total score were quite high, at .81 and .85, respectively. Test-retest correlations for SCWT interference score, WCST perseverative errors, and TMT time score were moderate, at .61, .50, and .49, respectively. To summarize, performance increased significantly from Time 1 to Time 2 for two of the five EF measures, and test-retest correlations for all five measures were in the moderate-to-high range.

Exploratory Factor Analysis

Exploratory factor analyses were then performed in order to examine the similarity of the factor structure for the EF measures at the two measurement occasions.

Step 1. The executive measures at Time 1 were entered into a principal axis factor analysis with Promax rotation. This analysis yielded a two-factor solution accounting for 60.9% of the variance (see Table 4). The first factor was weighted more heavily by WCST Perseverative Errors, TMT Composite Score, Category Fluency Total, and CFL Total than the second factor, and accounted for 40.6% of the variance. The second factor accounted for 20.3% of the variance, and was weighted by SCWT interference score more heavily than the first factor.

Step 2. All executive measures at Time 2 were entered into a principal axis factor analysis. This analysis yielded a one-factor solution accounting for 45.4% of the variance (see Table 4). Because only one factor was extracted, no rotation was necessary. SCWT interference score had a weaker loading on this factor than the other executive measures.

Step 3. Consistent with predictions regarding the factor structure of the executive measures, only one factor contributed meaningfully to interpretation of factors at both Time 1 and Time 2. Further exploratory analyses were therefore restricted to one-factor solutions. Principal axis factoring was conducted again with this restriction on all executive measures at Time 1 (see Table 5). The factor extracted accounted for 40.6% of the variance. SCWT interference score did not contribute meaningfully to this factor.

Step 4. Because SCWT interference score had a substantially lower factor loading than all other executive measures at both Time 1 and Time 2, this measure was excluded from further factor analyses. Principal axis factoring, restricted to one factor, was conducted again with all executive measures except for SCWT interference score, at Time 1 and Time 2, respectively (see Table 6). Removal of SCWT interference score from analyses increased the amount of variance accounted for at Time 1 to 50.8%, and at Time 2 to 54.4%.

To summarize, a one-factor model including four of the five original measures of executive function emerged from a series of exploratory factor analyses. The SCWT was excluded from this model because it lacked communality with the other executive measures.

Structural Analysis of Longitudinal Stability

Structural analyses were then performed on the four remaining executive measures in order to examine the longitudinal stability of executive function in this sample. Model O_1, the baseline model for these analyses, provided good overall model fit ($\chi^2 = 30.38$, df = 15, Bentler-Bonnet normed fit index (NFI) = .987). Model O_2, which constrained the factor loadings to be held constant across time, did not produce a significant change in model fit ($\chi^2 = 33.99$, df = 18, NFI = .986), and therefore the hypothesis that the factor loadings are invariant across time could not be rejected. Model O_3, which constrained the residual variance to be held constant across time, did not produce a significant change in model fit from model O_2 ($\chi^2 = 40.93$, df = 22, NFI = .983), and therefore the hypothesis that the residual variances are invariant across time could not be rejected. In this model, differences in the structure of EF from Time 1 to Time 2 are necessarily due to changes in the variance of EF itself, not in the degree to which each variable loads on EF or in the variance of measurement error. Model O_4, which constrained the factor variance of EF to be held constant across time, produced a significant decrease in fit from model O_3 ($\chi^2 = 45.80$, df = 23, NFI = .981), and therefore the hypothesis that the variance in EF is invariant across time was rejected. Variance of executive function for the longitudinal model is presented in Table 9. Model O_3 therefore became the accepted model of longitudinal EF factor structure. Goodness-of-fit indices for models O_1 through O_4 are presented in Table 6. To summarize, both the loadings of the executive measures on a single factor (EF) and the error variance associated with these measures were similar across measurement occasions. However, the variance of EF was greater at Time 2 than at Time 1.

Next, structural analyses were performed in order to examine the similarity of the effect of age on executive function at the two measurement occasions. Model O_5, the baseline model for structural analysis, allowed age to load freely on EF at both time points, was otherwise similar to the accepted model of longitudinal EF factor structure (model O_3), and provided good overall model fit ($\chi^2 = 53.80$, df = 28, NFI = .982). Standardized regression weights of age on EF_1 and EF_2 were estimated to be .50 and .60, respectively. Model O_6, which constrained the loadings of age on EF to be constant across time, produced a significant decrease in fit from model O_5 ($\chi^2 = 63.75$, df = 29, NFI = .979), and therefore the hypothesis that the effect of age on EF is invariant across time was rejected. Model O_5 therefore became the accepted structural model of the longitudinal effects of age on EF. Goodness-of-fit indices for models O_5 and O_6 are presented in Table 6. To summarize, the effect of age on executive function increased from the first to the second measurement occasion.

DISCUSSION

The purpose of this study was to investigate the unity and stability of executive function in an aging population, as measured by five widely-administered neuropsychological tests. Overall, the results suggest that executive function is a unified and stable construct. First, test-retest correlations for all five measures were in the moderate-to high range. Second, exploratory factor analyses and structural analyses revealed a high degree of unity and stability in the construct across time. Independent exploratory factor analyses of these measures of executive function at two measurement occasions yielded similar one-factor models that accounted for more than 50% of the observed variance. These models included WCST, TMT, CFL, and CAT tasks; SCWT was dropped from analysis after demonstrating little communality with the other measures. The degree of convergence demonstrated between the five executive measures in this study is remarkable, especially when the relatively heterogeneous nature of these tasks is taken into account.

Further analyses of the short-term longitudinal stability of this model of executive function supported the notion that executive function is a stable cognitive construct that can be measured reliably with commonly-used clinical measures. In fact, the only element of the longitudinal model that was found to be significantly invariant across the two measurement occasions was the variance of the latent executive function variables themselves.

Several elements separate this study from those that have been conducted previously, and may account for some of the results. First of all, no known studies have previously examined the longitudinal stability of the latent executive construct, as was done in this study. Secondly, none of the studies surveyed used the same group of executive measures that were found in this study (see Burgess et al., 1998; Duncan, Johnson, Swales & Freer, 1997; Lamar, Zonderman, & Resnick, 2002; Della Sala, Gray, Spinnler, and Trivelli, 1998; Boone et al., 1998; Miyake et al., 2000). It is reasonable to expect that not all executive measures would exhibit the degree of convergence that has been demonstrated in this study. Specifically, those measures that place substantial non-executive demands upon participants are likely to share less commonality than more cognitively "pure" executive measures (Miyake et al., 2000).

Although virtually every psychological measure inherently contains both executive and non-executive task demands, an effort was made in this study to include indices of executive function that were theoretically and statistically freer from contamination from non-executive processes (Miyake et al., 2000). For example, this study used a (Part B – Part A) algorithm in order to remove some of the effects of processing speed and visual search abilities. In addition, number of perseverative errors on the WCST was used rather than one or more of the many other calculable indices of the WCST, based upon the established relationship between executive dysfunction and the tendency to perseverate (Stern & Prohaska, 1996). This contrasts sharply with the factor analysis conducted by Lamar, Zonderman, and Resnick (2002), which analyzed TMT Part A and Part B simultaneously, and the factor analysis conducted by Boone, et al. (1998), which included four indices from WCST, and resulted in the extraction of a WCST "factor".

The relatively high reliability of the executive measures that were used may have also contributed to the surprising convergence demonstrated in this study. The average test-retest correlation for the five executive measures used was .65. Contrary to the notion that executive tasks can be used effectively only once (see Lowe & Rabbitt, 1998; Rabbitt, 1997; Phillips, 1997), these results suggest that many executive tasks may be reliable across at least two time points, even after relatively short delays (four to eight weeks). Even more impressive is the correlation in the accepted longitudinal model between the latent executive function variables at Time 1 and Time 2 of .96. The strength of this correlation suggests that an individual's overall picture of executive functioning, as derived from the four measures discussed here, is remarkably robust, and unlikely to change significantly over a short period of time.

There were, however, several unexpected results. The most important of these is the lack of communality demonstrated between the SCWT and other executive measures. It is not possible at the present time to compare this result with results that have been obtained previously, because none of the factor analyses surveyed included SCWT interference score as a variable of interest. While no label more specific than 'executive function' could be meaningfully applied collectively to the WCST, TMT, CFL, and CAT, which did demonstrate a high degree of communality, it is nevertheless possible that these measures all fall within some subset of executive function, and that the SCWT lies within a separate subset. Regardless, this result is informative in that it reduces the temptation to conclude from this study that all well-validated measures of executive function are likely to converge in all populations.

Additionally, it is surprising that the variance of the latent executive function variable at Time 2 was significantly greater than the variance at Time 1. It was presumed that if the variance of executive function were to change within the span of four to eight weeks, it would decrease due to the overall restriction of variance that often accompanies practice effects. Instead, it is possible that practice increased the performance of some subset of the sample, whereas the performance of others remained relatively unchanged. This type of differential practice effect would cause the overall variance in executive function to increase.

The hypothesis that the effect of age on executive function would be longitudinally stable was also rejected. This result suggests that age could be the basis for the potential differential in practice effects discussed above. If this were true, it would indicate that younger individuals in this study developed tasks skills to a greater extent than older individuals. Although it is important to note these possibilities, it is also important to note that the standardized regression weights of age on executive function at Time 1 (.50) and Time 2 (.60) are not dramatically different, and that the chi-square test of significance was based on a change in one degree of freedom. Indeed, the increase in executive function variance discussed above may have been sufficient to cause a significant increase in the relationship between executive function and age. Likewise, although the difference in the Bentler-Bonnet normed fit index changed by only .003, this difference was statistically significant because of the high sensitivity of the chi-square statistic.

Several aspects of the present study place limitations on the conclusions that can be drawn from the results. Because this study was conducted on a sample of non-demented, predominantly Caucasian older individuals with relatively high levels of education, some caution must be exercised when generalizing these results to the aging population at large. Similarly, the conclusions drawn in this study may not apply to other populations commonly examined by this research paradigm, such as college students or neurological patients. Finally, the results of this study are limited by the possibility that the five measures of executive function employed in this study may not be representative of the hundreds of measures used in worldwide psychological research and clinical practice.

Further investigations of executive function with different measures, and in different populations, would help clarify the degree to which the unity, reliability, and stability of executive function found in this study can be generalized. Expanding the number of measurement occasions to three or more, over longer periods of time, would also be beneficial in that it would provide a broader picture of the stability in executive function. Finally, the inclusion of brain imaging techniques within this research paradigm would also be helpful in clarifying the functional processes underlying both the unity and diversity in executive function.

In conclusion, the results of this study suggest that executive function is a relatively unified, stable construct. These findings challenge views of executive function as a highly fractionated construct (see Burgess et al., 1998; Duncan, Johnson, Swales, & Freer, 1997; Lamar, Zonderman, and Resnick, 2002) that cannot be measured reliably more than once (see Lowe & Rabbitt, 1998; Rabbitt, 1997; Phillips, 1997). However, considering the limitations of this study, clarifications of these results might be achieved through additional research. In particular, designs that include more executive tasks and more measurement occasions, over longer periods of time, and in a variety of other samples, are likely to be particularly valuable.

APPENDIX

Table 1

Demographic Characteristics of Sample

Age	54-58	59-63	64-68	69-73	74-78	79-83	84-87	88+	N Missing	Total
N	17	23	25	29	33	14	7	1	2	151
% Female	70.6	69.6	68.0	51.7	54.5	71.4	14.3	100.0	0	60.3
Mean Years Education	16.06	15.43	16.04	15.69	15.48	14.93	15.00	14.00	1	15.59

Table 2

Descriptive Statistics of Executive Variables at Time 1 and Time 2

Variable	Valid N	Mean[a] (Time 1)	Mean[a] (Time 2)	Mean Difference[b]	p
WCST[c]	109	17.38 (15.27)	14.39 (13.50)	-2.98	< .05
TMT[d]	119	59.87 (40.80)	62.99 (48.33)	3.13	*ns*
CFL[e]	121	38.20 (11.42)	39.64 (12.01)	1.44	< .05
CAT[f]	120	44.58 (10.85)	45.37 (11.65)	.79	*ns*
SCWT[g]	116	-4.29 (7.04)	-3.58 (8.07)	.71	*ns*

[a] Standard deviations in parentheses

[b] Calculated as (Time 2 – Time 1); missing data excluded pairwise

[c] Wisconsin Card Sorting Test perseverative errors

[d] Trail Making Test time score (Part B – Part A)

[e] Letter Fluency total score

[f] Category Fluency total score

[g] Stroop Test interference score

Table 3

Test-Retest Correlations

Variable	r_{12} [a]
WCST[b]	.50
TMT[c]	.49
CFL[d]	.81
CAT[e]	.85
Stroop[f]	.61

[a] Variables autocorrelated at two time points
[b] Wisconsin Card Sorting Test perseverative errors
[c] Trail Making Test time score (Part B – Part A)
[d] Letter Fluency total score
[e] Category Fluency total score
[f] Stroop Interference Score

Table 4

Correlations and Factor Loadings of Executive Functioning Measures, Eigenvalue > 1 Method

Time 1	1	2	3	4	5	Factor 1	Factor 2
1. WCST	--	.27**	-.03	-.29**	-.22*	-.40	.16
2. TMT		--	-.06	-.36**	-.38**	-.58	.13
3. SCWT			--	.10	-.02	.10	-.26
4. CAT				--	.50**	.74	-.18
5. CFL					--	.69	.29
Eigenvalue						2.03	1.02
% of variance						40.56	20.33

Time 2	1	2	3	4	5	Factor 1
1. WCST	--	.53**	-.13	-.39**	-.18*	-.53
2. TMT		--	-.29**	-.49**	-.30**	-.76
3. SCWT			--	.22*	.03	.31
4. CAT				--	.47**	.74
5. CFL					--	.44
Eigenvalue						2.27
% of variance						45.44

Table 5

Correlations and Factor Loadings of Executive Functioning Measures, One Factor Extracted

	Time 1	Time 2
1. WCST	-.39	-.53
2. TMT	-.58	-.76
3. SCWT	.07	.31
4. CAT	.73	.74
5. CFL	.64	.44
Eigenvalue	2.03	2.27
% of variance	40.56	45.44

Table 6

Correlations and Factor Loadings of Executive Functioning Measures, One Factor Extracted and SCWT Excluded

	Time 1	Time 2
1. WCST	-.40	-.59
2. TMT	-.59	-.75
4. CAT	.71	.71
5. CFL	.65	.46
Eigenvalue	2.03	2.17
% of variance	50.82	54.35

Table 7

Goodness-of-Fit Statistics for Alternative Models

Model	χ^2	*df*	*p*	ρ^a	Comparison	$\Delta\chi^2$	Δdf	*p*	$\Delta\rho^a$
O_1	30.38	15	.011	.987	---	---	---	---	---
O_2	33.99	18	.013	.986	$O_1 - O_2$	3.61	3	*ns*	.001
O_3	40.93	22	.008	.983	$O_2 - O_3$	6.95	4	*ns*	.003
O_4	45.80	23	.003	.981	$O_3 - O_4$	4.87	1	$< .05$	.002
O_5	53.80	28	.002	.982	---	---	---	---	---
O_6	63.75	29	.000	.979	$O_5 - O_6$	9.95	1	$< .01$	.003

Note. n = 151 for all analyses.

[a] Bentler-Bonnet normed fit index.

Table 8

Factor Loadings and Residual Variances for the Longitudinal Factor Model (O_3)

Variable	AMOS estimates[a]	Standardized Loadings Time 1	Time 2	Residual Variances[b]	Residual Covariances	Residual Correlations
WCST[c]	1.00[d] (--)	.46	.51	144.76	48.04	.33
TMT[e]	3.99 (.741)	.59	.64	1193.40	275.76	.23
CFL[f]	-0.98 (.200)	-.52	-.58	102.30	77.45	.76
CAT[g]	-1.22 (.223)	-.71	-.76	57.17	41.46	.73

[a] Standard errors in parentheses

[b] Calculated as the proportion of residual variance to total variance; 1 – (residual variance) = communality

[c] Wisconsin Card Sorting Test perseverative errors

[d] Fixed Parameter

[e] Trail Making Test time score (Part B – Part A)

[f] Letter Fluency total score

[g] Category Fluency total score

Table 9

Variance of Executive Function

for the Longitudinal Factor Model (O_3)

Variance	
Time 1	39.79 (13.02)
Time 2	52.48 (16.86)
Covariance	43.74 (14.05)
Correlation	.96

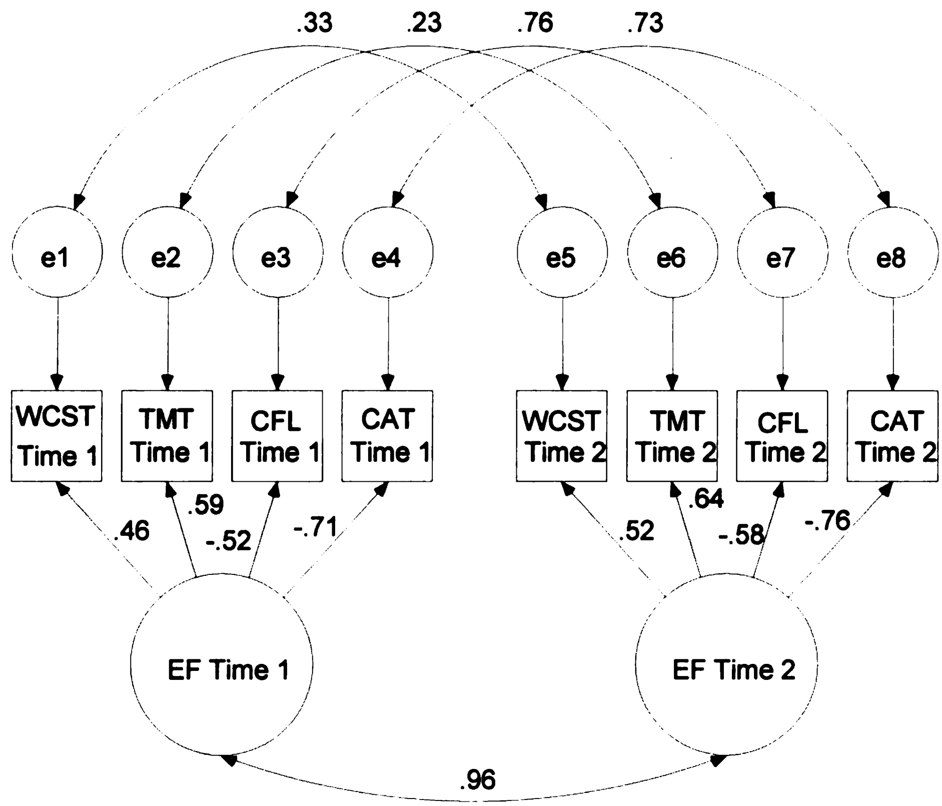

Figure 1. Accepted model of longitudinal EF factor structure (Model O_3).
Regression weights and residual variances of executive measures have been held constant across time.
Standardized values shown.

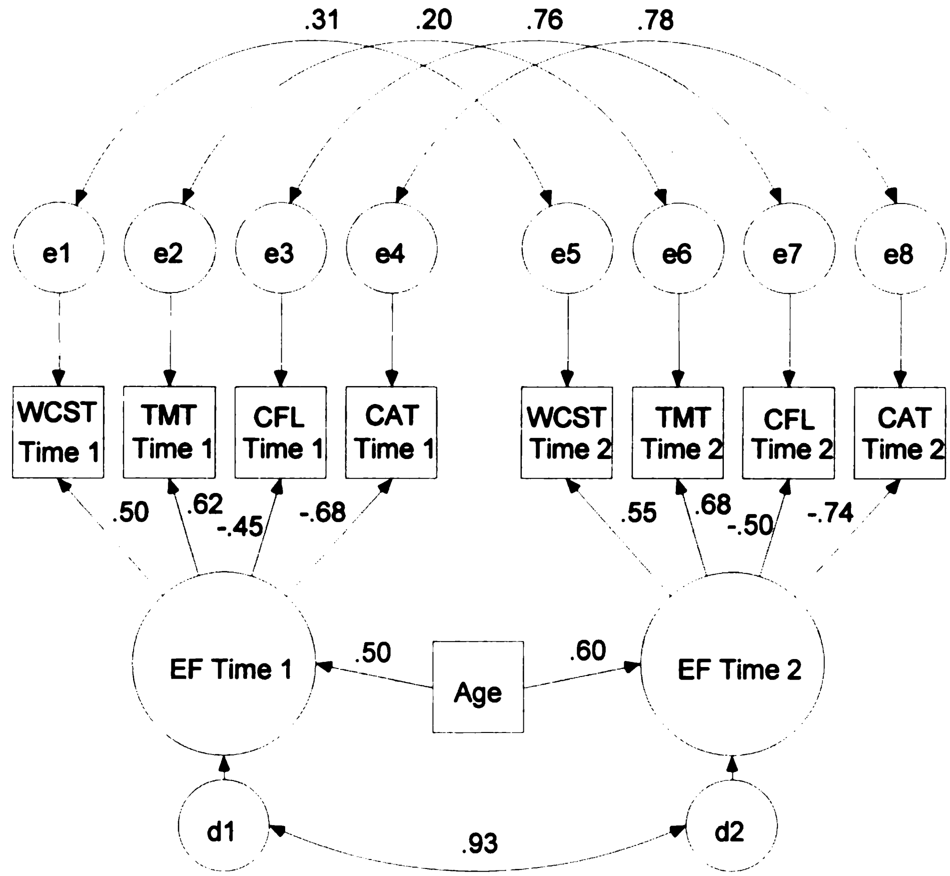

Figure 2. Accepted model of relationship of age to longitudinal EF factor structure (Model O_5). Regression weights and residual variances of executive measures have been held constant across time. Standardized values shown.

References

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 & Experimental Neuropsychology, 13(6), 909-922.

Arbuthnott, K., & Frank, J. (2000). Trail Making Test, part B as a measure of executive control: Validation using a set-switching paradigm. Journal of Clinical and Experimental Neuropsychology, 22(4), 518-528.

Axelrod, B. N., Goldman, R. S., & Henry, R. (1992). Sensitivity of the Mini-Mental State Examination to frontal lobe dysfunction in normal aging. Journal of Clinical Psychology, 48(1), 68-71.

Band, G. P., Ridderinkhof, K. R., & Segalowitz, S. (2002). Explaining neurocognitive aging: is one factor enough? Brain and Cognition, 49, 259-267.

Barcelo, F. (2001). Does the Wisconsin Card Sorting Test measure prefrontal function? The Spanish Journal of Psychology, 4(1), 79-100.

Beck, A. T., Steer, R. A., Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical Psychology Review, 8(1), 77-100.

Beck, A.T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1983). Multilingual aphasia examination (3rd ed.). Iowa City, IA: AJA Associates.

Berman, K. F., Ostrem, J. L., Randolph, C., Gold, J., Goldberg, T. E., Coppola, R., Carson, R. E., Herscovitch, P., & Weinberger, D. R. (1995). Physiological activation of a cortical network during performance of the Wisconsin Card Sorting Test: A positron emission tomography study. Neuropsychologia, 33(8), 1027-1046.

Bilger, E. D. (1997). Neuroimaging in normal aging and dementia. In Nussbaum, P. D. (Ed.), Handbook of Neuropsychology and Aging (pp. 409-421). New York: Plenum Press.

Boone, K. B., Ponton, 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(7), 585-595.

Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: considerations for detecting adult age differences. Journal of Clinical and Experimental Neuropsychology, 22(1), 40-55.

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). Cognitive neuroscience of aging: contributions of functional neuroimaging. Scandinavian Journal of Psychology, 42, 277-286.

Celine, S., Michel, I., & Espagnet, L. (2000). Aging, episodic memory Feeling-of-Knowing, and frontal functioning. Neuropsychology, 14(2), 299-309.

Craik, F., Morris, L. W., Morris, R. G., & Loewen, E. R. (1990). Relations between source amnesia and frontal lobe functioning in older adults. Psychology and Aging, 5(1), 148-151.

Crowe, S. F. (1998). The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. Journal of Clinical Psychology, 54(5), 585-591.

Della Sala, S., Gray, C., Spinnler, H., & Trivelli, C. (1998). Frontal lobe functioning in man: the riddle revisited. Archives of Clinical Neuropsychology, 13(8), 663-682.

Duncan, J., Emslie, H., & Williams, P. (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 lobe deficits after head injury: unity and diversity of function. Cognitive Neuropsychology, 14(5), 713-741.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4(3), 272-299.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12(3), 189-198.

Foster, J. R., Sclan, S., Welkowitz, J., Boksay, I., & Seeland, I. (1988). Psychiatric assessment in medical long-term care facilities: reliability of commonly used rating scales. International Journal of Geriatric Psychiatry, 3(3), 229-233.

Frith, C. D., Friston, K. J., Liddle, P. F., Frackowiak, R. S. (1991). A PET study of word finding. Neuropsychologia, 29(12), 1137-1148.

Fuster, J. M. (1997). The Prefrontal Cortex (3rd ed.). Philadelphia, PA: Lippincott-Raven Publishers.

Garden, S. E., Phillips, L. H., & MacPherson, S. E. (2001). Midlife aging, open-ended planning, and laboratory measures of executive function. Neuropsychology, 15(4), 472-482.

Giaquinto, S. (1988). Aging and the Nervous System. New York: John Wiley and Sons.

Goldberg, E. (2001). The Executive Brain. New York: Oxford University Press.

Gourovitch, M. L., Kirkby, B. S., Goldberg, T. E., Weinberger, D. R., Gold, J. M., Esposito, G., Van Horn, J. D., & Berman, K. F. (2000). A comparison of rCBF patterns during letter and semantic fluency. Neuropsychology, 14(3), 353-360.

Gur, R. C., Gur, R. E., Orbist, W. D., Skolnick, B. E., & Reivich, M. (1987). Age and regional cerebral blood flow at rest and during cognitive activity. Archives of General Psychiatry, 44, 617-621.

Haug, H., & Eggers, R. (1991). Morphometry of the human cortex cerebri and corpus striatum during aging. Neurobiology of aging, 12(4), 336-338.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). Wisconsin Card Sorting Test Manual. USA: Publishing Assessment Resources.

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. Psychology and Aging, 1(2), 159-171.

Konishi, S., Kawazu, M., Uchida, I., Kikyo, H., Asakura, I., & Miyashita, Y. (1999). Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. Cerebral Cortex, 9(7), 745-753.

Lamar, M., Zonderman, A. B., & Resnick, S. (2002). Contribution of specific cognitive processes to executive functioning in an aging population. Neuropsychology, 16(2), 165-162.

LaRue, A. (1992). Aging and Neuropsychological Assessment. New York: Plenum Press.

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

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., & Hommer, D. W. (1999). Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. Journal of Clinical and Experimental Neuropsychology, 21(1), 2-16.

Lowe, C., & Rabbitt, P. (1998). Test/retest reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Neuropsychologia, 36(9), 915-923.

Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., & Cohen, N. J. (2002). Attentional control in the aging brain: insights from an fMRI study of the Stroop task. Brain and Cognition, 49, 277-296.

Mitrushina, M., & Satz, P. (1991). Effect of repeated administration of a neuropsychological battery in the elderly. Journal of Clinical Psychology, 47(6), 790-801.

Mitrushina, M., Uchiyama, C., & Satz, P. (1995). Heterogeneity of cognitive profiles in normal aging: Implications for early manifestations of Alzheimer's Disease. Journal of Clinical and Experimental Neuropsychology, 17(3), 374-382.

Miyake, A., Friedman, N., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cognitive Psychology, 41, 49-100.

Phillips, L. H. (1997). Do "frontal tests" measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbit (Ed.), Methodology of Frontal and Executive Function (pp. 191-213). Psychology Press: East Sussex, U.K.

Rabbitt, P. (1997). Methodology of Frontal and Executive Function. Psychology Press: East Sussex, U.K.

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), 95-114.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. Perceptual & Motor Skills, 8, 271-276.

Rezai, K., Andreasen, N. C., Alliger, R., Cohen, G., Swayze, V., & O'Leary, D. S. (1993). The neuropsychology of the prefrontal cortex. Archives of Neurology, 50(6), 636-642.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. M. (1998). A study of performance on tests from the

CANTAB battery sensitive to frontal lobe dysfunction. Journal of the International Neuropsychological Society, 4, 474-490.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. Journal of Cognitive Neuroscience, 12(1), 142-162.

Rympa, B., Prabhakaren, V., Desmond, J. E., & Gabrieli, J. D. (2001). Age differences in prefrontal cortical activity in working memory. Psychology and Aging, 16(3), 371-384.

Salthouse, T. A. (1991). Theoretical Perspectives on Cognitive Aging. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? Neuropsychology, 10(2), 272-285.

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., & Barta, P. (2000). Elucidating the contributions 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.

Seeman, T. E., Lusignolo, T. M., Albert, M., & Berkman, L. (2001). Social relationships, social support, and patterns of cognitive aging in healthy, high-functioning older adults: MacArthur Studies of Successful Aging. Health Psychology, 20(4), 243-255.

Shaw, T. G., Mortel, K. F., Meyer, J. S., Rogers, R. L., Hardenburg, J., & Cutaia, M. M. (1984). Cerebral blood flow changes in benign aging and cerebrovascular disease. Neurology, 34, 855-862.

Silver, M. H., & Perls, T. T. (2001). Is dementia the price of a long life? An optimistic report from centenarians. Journal of Aging and Health, 13(2), 267-286.

Silverman, P., Hecht, L., & McMillin, D. (2000). Modeling life satisfaction among the aged: A comparison of Chinese and Americans. Journal of Cross-Cultural Gerontology, 15, 289-305.

Spreen, O., & Strauss, E. (1998). A compendium of neuropsychological tests : administration, norms, and commentary. New York : Oxford University Press.

Stern, R. A., & Prohaska, M. L. (1996). Neuropsychological evaluation of executive functioning. American Psychiatric Press Review of Psychiatry, 15, 243-266.

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

Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. Neuropsychologia, 39, 771-786.

Vendrell, P., Junque, C., Pujol, J., Jurado, M. A., Molet, J., & Grafman, J. (1995). The role of prefrontal lesions in the Stroop task. Neuropsychologia, 33(3), 341-352.

Wang, L., Kakigi, R., & Hoshiyama, M. (2001). Neural activities during Wisconsin Card Sorting Test – MEG observation. Cognitive Brain Research, 12, 19-31.

Wecker, S. W., Kramer, J.H., Wisniewski, A., Delis, D. C., & Kaplan, E. (2000). Age effects on executive ability. Neuropsychology, 14(3), 409-414.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. Psychological Bulletin, 120(2), 272-292.

West, R. L., & Bell, M. A. (1997). Stroop color--word interference and electroencephalogram activation: Evidence for age-related decline of the anterior attention system. Neuropsychology, 11(3), 421-427.

Winocur, G., Moscovitch, M., Stuss, D. T. (1996). Explicit and implicit memory in the elderly: Evidence for double dissociation involving medial temporal- and frontal-lobe functions. Neuropsychology 10(1), 57-65.

Yesavage, J. A., Brink, T.L., Rose, T.L., Lum O., Huang, V., Adey, M., & Leirer, V.O. (1983). Development and validation of a Geriatric Depression Screening Scale: A preliminary report. Journal of Psychiatric Research, 17(1), 37-49.

Ylikoski, R., Ylikoski, A., Keskivaara, P, Tilvis, R., Slkava, R., & Erkinjuntti, T. (1999). Heterogeneity of cognitive profiles in aging: successful aging, normal aging, and individuals at risk for cognitive decline. European Journal of Neurology, 6, 645-652.