Published in final edited form as:

Neuropsychology. 2009 November ; 23(6): 778–791. doi:10.1037/a0016743.

Characterizing Executive Functioning in Older Special Populations: From Cognitively Elite to Cognitively Impaired

Cindy M. de Frias,

Department of Psychology, Stockholm University

Roger A. Dixon, and

Department of Psychology, University of Alberta

Esther Strauss*

Department of Psychology, University of Victoria

Abstract

The authors examined the structure and invariance of executive functions (EF) across (a) a continuum of cognitive status in 3 groups of older adults (cognitively elite [CE], cognitively normal [CN], and cognitively impaired [CI]) and (b) a 3-year longitudinal interval. Using latent variable analyses (LIS-REL 8.80), the authors tested 3-factor models (“Inhibition”: Hayling [Burgess & Shallice, 1997], Stroop [Regard, 1981]; “Shifting”: Brixton [Burgess & Shallice, 1997], Color Trails [D’Elia et al., 1996]; and “Updating”: Reading and Computational Span [Salthouse & Babcock, 1991]) and 1-factor models within each group. Participants (initial $N=570$; 53–90 years) were from the Victoria Longitudinal Study (Sample 3, Waves 1 and 2). Cross-sectionally, the authors observed a 3-factor EF structure especially for the CE group and 1-factor solutions for all 3 groups. Longitudinally, temporal invariance was supported for the 3-factor model (CE and CN groups) and the 1-factor model (CI and CN groups). Subgroups with higher cognitive status and greater 3-year stability performed better on EF factors than corresponding groups with lower cognitive status and less stability. Studies of EF structure, performance, dedifferentiation, and dysfunction will benefit from considering initial cognitive status and longitudinal stability.

Keywords

executive functions; cognitive status; measurement invariance; aging; Victoria Longitudinal Study

As adults engage in both routine and novel cognitive situations, a set of executive functions (EF) serve to control and promote effective operations. Such control can occur through processes that reflect the initiation, regulation, planning, monitoring, maintaining, stopping, and altering of specific behavior (Stuss, Shallice, Alexander, & Picton, 1994). Several possible underlying aspects or factors of EF have been identified conceptually and tested empirically using psychometric (e.g., correlations with other measures, factor analysis) and clinical (e.g., special population) studies. Regarding the latter, a widening variety of populations have recently have been examined, including presumably neurologically intact young adults (Friedman & Miyake, 2004; Miyake et al., 2000), relatively healthy older adults (de Frias, Dixon, & Strauss, 2006; Ettenhofer, Hambrick, & Abeles, 2006), and

© 2009 American Psychological Association

Correspondence concerning this article should be addressed to Cindy M. de Frias, Frescati Hagväg 14, Psychology Department, Stockholm University, 106 91 Stockholm, Sweden. cdefrias@psychology.su.se.

*We regret that Esther Strauss passed away after final acceptance of this article.

various patient groups with disorders such as schizophrenia and depression (e.g., Marczewski, Van der Linden, & Laroi, 2001; Moritz et al., 2002; Smitherman, Huerkamp, Miller, Houle, & O'Jile, 2007). One influential perspective, supported by latent variable analyses, is that executive functioning involves at least three key processes: inhibition (e.g., restraint), shifting (e.g., switching), and updating (e.g., working memory; see Miyake et al., 2000).

Although a potentially multifaceted construct, firm conclusions about the independence and structure of EF have been limited by several measurement and assessment issues (e.g., Bielak, Mansueti, Strauss, & Dixon, 2006; Friedman et al., 2006; Salthouse, Atkinson, & Berish, 2003). First, multiple measures (varying in psychometric and normative characteristics) are used across studies and populations, and these measures often reflect different levels of analysis and degrees of process purity (cf. cognitive tasks, neuropsychological tests). Second, issues of conceptual independence between cognitive domains (e.g., intelligence, working memory) have been inconsistently identified, again across several sample compositions. Third, theoretically and clinically important variations in structure and performance that depend on group composition and populations studied (e.g., those with and without neurodegenerative diseases, disturbances in frontal lobes, or cognitive deficits) have been inferred or reported.

Commentaries and research on each of these issues have appeared in recent years (e.g., Burgess et al., 2006; Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003; Friedman et al., 2006; Lezak, Howieson, & Loring, 2004; Royall et al., 2002; Rabbitt, Lowe, & Shilling, 2001; Salthouse et al., 2003; Strauss, Sherman, & Spreen, 2006). In this article, we draw from a select subset of these issues, focusing on structural and temporal invariance of relationships among EF tasks in three clinically important groups of older adults. The present study is influenced by previous research on EFs and aging from several complementary perspectives. Some experimental work has shown that the construct of EF, as evaluated with available tasks and methods, may overlap with other constructs (e.g., fluid intelligence) in a manner that may qualify its promise as an independent explanatory mechanism to account for age-related cognitive differences (Luszcz & Lane, 2008; Salthouse et al., 2003). Some related psychometric research has used latent variable (covariance structure) methods to address the important challenges of (a) using clinically valuable neuropsychological tests of EF that can vary in their cognitive specificity (e.g., Miyake et al., 2000), (b) evaluating reliability and longitudinal stability of EF assessment (which has been discussed in the literature; Ettenhofer et al., 2006; Lowe & Rabbitt, 1998), and (c) testing structural and measurement invariance for at least some subsets of EF indicators in aging populations (e.g., de Frias et al., 2006; Hull, Martin, Beier, Lane, & Hamilton, 2008). For example, Ettenhofer et al. showed that for healthy older adults, EFs could be measured reliably (at the individual test level) and with stability (at the construct level) across a multiweek period (for perspectives see Lowe & Rabbitt, 1998; Luszcz & Lane, 2008). Considering actual performance characteristics, researchers from neuropsychological perspectives have reported selective impairments in various EFs across several aging-related neuropsychological conditions. For example, EFs are promising markers of cognitive deficits in mild cognitive impairment (MCI) and Alzheimer's disease (AD; Amieva, Phillips, Della Sala, & Henry, 2004; Rapp & Reischies, 2005; Ready, Ott, Grace, & Cahn-Weiner, 2003; Zhang, Han, Verhaeghen, & Nilsson, 2007). In one recent dissociation study based on latent variable analyses, Zhang et al. found MCI-related deficits in planning (Trail Making Test; Reitan & Wolfson, 1985) but not inhibition (Stroop [Regards, 1981]).

In an earlier study, we used covariance models to examine structural characteristics of four EF tests, including two relatively new and two standard tests, in a large sample ($n = 427$) of

healthy older adults from the Victoria Longitudinal Study (VLS; de Frias et al., 2006). Together, the four tests represented the inhibition (Hayling [Burgess & Shallice, 1997], Stroop) and shifting (Brixton [Burgess & Shallice, 1997], Color Trails [D'Elia, Satz, Uchiyama, & White, 1996]) aspects of EF. A single-factor model fit the data well, and was notably invariant across gender and two age groups (young-old and old-old). The sample included middle-aged and older adults (55 to 90 years old), so the comparison to potential multifactor solutions with younger adult populations was not possible. This limitation left open the conceivable clinical interpretation that dedifferentiation (from multidimensional to unidimensional representations) of EFs may have occurred over the adult life span, or that specific aging-related population groups may produce different solutions. In the present study, we explore these issues with several unique initiatives. First, we increased our sample size and tested it on a second longitudinal occasion (3 years after the first wave). These new VLS data permit analyses of change in level, but also potential change (or invariance) in structure, across time. The latter is a focus of the present article. Second, we added markers of a third EF aspect (i.e., updating through tests of working memory) to complement the measurement battery. Third, we used careful classification procedures to supplement our earlier study of normal aging. Specifically, in addition to a cognitively “normal” group, we used relative performances on a separate battery of cognitive reference measures to identify two additional groups. These are (a) a cognitively “elite” group that performs uniformly well on the cognitive reference measures and (b) a provisional mildly cognitively “impaired” group that performs substantially lower than the intermediate cognitively normal group on the battery.

In general, the overall goal is to explore structural and invariance questions across these three groups and over two waves (encompassing 3 years) of study. Although the precise mapping of structure to function is under development, changes in structure and level of performance in some executive function tests may be attributed to alterations in frontal-striatal circuits (Buckner, 2004). Such changes in nondemented older adults are variously attributed to (a) vascular disease (e.g., hypertension) that preferentially creates frontal white matter lesions (Gunning-Dixon & Raz, 2000), (b) atrophy of frontal gray matter (Raz et al., 1997), and (c) neural noise from decreased catecholamine levels that may degrade information processing (Li, Lindenberger, & Sikstrom, 2001). To the extent that these changes occur for all individuals in the same way (and at the same level) across adulthood an age-invariant unitary factor of EF may be expected. Alternatively, a differentiated factor structure, as described in healthy younger adults, may be observed for older adults to the extent that neurodegenerative processes occur differentially across brain structures and across aging individuals. Conceivably, adaptive features or mechanisms (such as cognitive reserve and neural and functional compensation) may prolong the onset of normative neurological decline (or permit cognitive manifestations consistent with this assumption). This could maintain differentiated structures and perhaps higher levels of performance, thereby delaying dedifferentiation of EFs (e.g., Buckner, 2004; Dixon, Garrett, & Bäckman, 2008; Reuter-Lorenz & Cappell, 2008; Stern, 2006).

In the present study, indirect evidence of dedifferentiation of EFs would take the form of a single-factor solution for the cognitively normal adults (i.e., replicating the results of de Frias et al., 2006) and a multifactor (e.g., three-aspect) solution for the cognitively elite adults (i.e., reflecting a potentially “maintained” multidimensional EF structure from young adulthood; e.g., Miyake et al., 2000). In all cases, further supportive evidence would be produced if the initial solutions at Wave 1 were to be sustained over 3 years (i.e., showing measurement invariance across time). Expectations regarding the impaired group are not certain, and range from unitary structures to multifactor solutions. Overall, regarding temporal invariance, we expect 3-year invariance of structures observed initially for each group. If observed, changes in the relations among the EF indicators over the 3-year period

could be attributed to several sources, including: (a) heterogeneity in the extent and timing of normative or impaired change (e.g., rate of decline or impairment may differ across individuals), and (b) unmeasured heterogeneity in group composition (e.g., as a function of comorbidities, disease etiology, or localization of impairment; see Sliwinski, Hofer, & Hall, 2003). If either of these processes is tapped in the 3-year longitudinal period of the present study, lack of temporal invariance may be indicated. Although not predicted, this outcome may be more likely with the cognitively normal and cognitively impaired groups, as the cognitively elite group membership requires more consistency and homogeneity of performance, and perhaps sustained neural integrity.

The existing literature supports differentiated executive functions in cognitively intact young adults and potential dedifferentiation in a sample of heterogeneous, typically aging older adults (i.e., unclassified on cognitive status), with virtually no information for older adults who are relatively cognitively elite or cognitively impaired. Using latent variable and other analyses, we addressed five goals concerning the organization of EFs across a continuum of cognitive status in three subgroups of older adults: cognitively elite (CE), cognitively normal (CN), and cognitively impaired (CI) adults. As appropriate our expectations for these goals are described above. The first goal was to test whether there is unity or diversity in the organization of six EF tasks in each of three groups (CE, CN, and CI). The second goal, based on the results of the first, was to test the multigroup (CE, CN, and CI) measurement invariance of the structure of EFs. The third goal was to test the longitudinal invariance of the previously identified structure of EFs within each cognitive status group. The fourth goal was to explore whether stability in classification affects the structure and invariance of EFs in the three groups. The fifth goal tested whether 3-year stability or nonstability in group classification affected level of EF performance.

Method

Participants

The participants were community-dwelling adults from the VLS, an ongoing multisample sequential study of cognitive, neuropsychological, health, and biological aging. They were originally recruited through advertisements in the public media and through appeals to community groups. They were paid nominal fees for their participation. The longitudinal data used for this study were from the first and second waves of testing (a 3-year interval) for the third sample (VLS Sample 3). From the VLS neuropsychological battery, six tasks measuring executive functioning were included. At Wave 1 (W1), $n = 570$ participants (385 women and 185 men) who were aged 53 to 90 years ($M_{age} = 68.2$ years, $SD = 8.6$ years) were tested. The average level of education was 15.2 years ($SD = 3.0$ years). At Wave 2 (W2), $n = 399$ participants (270 woman and 129 men) ($M_{age} = 72.2$ years, $SD = 8.5$ years) were available. The average level of education was 15.6 years ($SD = 2.9$ years). All participants attained scores above 23 on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). Self-reported health was referenced by two indicators: absolute health (i.e., health relative to a perfect state) and relative health (i.e., health relative to same-age peers). For both indicators, the participants from each group rated their health as being good to very good on a 5-point scale, in which 1 indicated very good health (see below and Table 1). Published overviews present further background information on the VLS methods, procedures, and samples (Dixon & de Frias, 2004; Hultsch, Hertzog, Dixon, & Small, 1998).

Participants at W1 were classified into three groups representing a continuum of provisional cognitive status: CE, CN, and CI. Cognitive status classification was determined by a procedure adapted from previous VLS and other research (e.g., Dixon et al., 2007; Ritchie, Artero, & Touchon, 2001; Tuokko & Hultsch, 2006). First, prior to classifying persons as

CE, CN, or CI, the participants were stratified by age (53 to 70 and 71 to 90 years) and level of education (0 to 12 years or 13 + years), and placed into one of four groups. Young-old (YO; low education) $n = 58$; YO (high education) $n = 292$; Old-old (OO; low education) $n = 49$; OO (high education) $n = 171$. Second, for each of the four groups, mean performance was calculated for five cognitive reference tests, representing the cognitive domains of perceptual speed, inductive reasoning, episodic memory, verbal fluency, and semantic memory. The resulting distributions served as within-sample norms for cognitive status classification (Dixon et al., 2007). Third, based on W1 scores, we classified participants into one of three mutually exclusive (but conceptually continuous) cognitive status groups. To summarize the logic, we first adopted the objective (cognitive reference test) procedures used in related MCI literature (e.g., Dixon et al., 2007) for defining our CI group. Second, most previous MCI research classifies an impaired group and clusters the remainder into a normal control group. We subdivided the latter group by modifying the selection procedures to reflect consistently higher level performance for CE participants. Third, we defined the intermediate group (scoring neither too low or too high) as CN. Missing data on executive measures reduced the sample size at W1 ($n = 501$) and W2 ($n = 341$). The group composition below (and information in Table 1) reflects this adjustment. First, all members of the CE group (W1 $n = 80$, W2 $n = 48$) scored above the mean on each of the five reference tests at W1. Second, the members of the CN group (W1 $n = 284$, W2 $n = 202$) scored between $-1.5 SD$ and $+1.5 SD$ on all five reference tests, with at least one score below the relevant group mean. All members of the CI group (W1 $n = 137$, W2 $n = 91$) produced at least one score that was $<-1.5 SD$ below the group mean. Descriptive information about the participants at W1 and stratified by cognitive status is reported in Table 1.

Executive Function Measures

We used six cognitive neuropsychological tests to measure EFs representing three key aspects, namely inhibition, shifting, and updating. The first four tests listed below were used in our previous study (de Frias et al., 2006) and tap inhibition (Hayling, Stroop) and shifting (Brixton, Color Trails 2), whereas the two working memory tests (Computational Span, Reading Span) are included to reflect updating. Psychometric properties of these tests have been reported and all have been used in multiple studies, including healthy aging and clinical populations (Strauss et al., 2006).

Hayling sentence completion test—This test (Burgess & Shallice, 1997) consists of two sets of 15 sentences, each having the last word missing. In the first part, Section 1, the examiner reads each sentence aloud and the participant has to complete the sentence as quickly as possible. For example, “the old house will be torn ... [participant says] down.” This section yields a simple measure of response speed. In the second part, Section 2, the participant completes the sentence with a word that is unconnected to the sentence in every way. For example, “the captain wanted to stay with the sinking ... [participant says] light bulb.” Therefore, the participant has to inhibit (restrain) a strongly activated (automatic) response before generating a new response. Completing the sentence with a word that is connected to the sentence in meaning would be coded as an error (i.e., the inability to inhibit an automatic response). The potential processes involved in the Hayling test are inhibition and initiation. This task yields three scores, an error score (from Section 2) and two measures of response speed (from Sections 1 and 2). Scaled scores (range = 1 to 10) were derived from cutoffs of the raw score latencies and errors. Using the method described in the test manual, an overall scaled score was derived as follows: scaled score Section 1—scaled score Section 2—errors Section 2, and this score based on a scale ranging from 1 (*impaired*) to 10 (*very superior*) was used for this analysis.

Stroop test—This task is thought to tap inhibitory processes by requiring the respondent to ignore the automatic response of reading a printed word (attending to verbal content) and to instead name the color of ink in which the word is printed (Regard, 1981). In Part A, the participant names as fast as possible the color of 24 dots printed in blue, green, red, or yellow. Each color is used six times, and the four colors are arranged in a pseudorandom order within the array, with each color appearing once in each row. Part B is similar to Part A except that the dots are replaced by common (noncolor) words (e.g., when, hard, and over), printed in lower case. The respondent is instructed to name the colors in which the word is printed and to ignore their verbal content. In Part C, the colored stimuli are the color names (i.e., blue, green, red, and yellow) printed in lower case with the color being incongruent to the color name. The performance score was the interference index $([Part\ C - Part\ A]/Part\ A)$.

Brixton test—The Brixton Spatial Anticipation Test (Burgess & Shallice, 1997) is a rule-attainment (or shifting) task based on the Wisconsin Card Sorting Task (WCST; Psychological Assessment Resources, 1990). However, in contrast to the WCST, the rule that is currently in operation cannot be triggered by any perceptually salient aspect of the stimuli. The Brixton test consists of a 56-page stimulus booklet, with each page showing the same basic array of 10 circles set in two rows of 5, with each circle numbered from 1 to 10. On each page, one of the circles is filled with a blue color. The position of this filled circle changes from one page to the next. The changes in position are governed by a series of simple rules that vary without warning. The participant is shown one page at a time and is asked to decide where the next filled position will be, basing the selection on an apparent pattern or rule derived from the previous pages. The test does not require verbal responses; rather, the participant can point to a position. Using the method described in the test manual, the total errors were recorded and these errors (maximum 54) were converted to scaled scores (e.g., 0 to 7 raw errors were converted to a scaled score of 10, which is classified as *very superior performance*). An overall standardized scaled score based on a scale ranging from 1 (*impaired*) to 10 (*very superior*) was used for analysis.

Color Trials Test (CTT)—The CTT (D'Elia et al., 1996) is similar to the Trial Making Test (Reitan & Wolfson, 1985) but was designed to minimize the influence of language. The potential executive process involved in this task is shifting. Part 1 requires the participant to connect, by making pencil lines, encircled numbers 1 through 25 randomly arranged on a page, in proper order. Odd numbered circles have a pink background, and all even-numbered circles have a yellow background. Part 2 (CTT-2) shows all numbers from 1 to 25 twice (one sequence with a yellow background, the other in pink). The participant is required to connect the numbers from 1 to 25 alternating between pink and yellow circles and disregarding the numbers in circles of the alternate color (D'Elia et al., 1996). The time for the completion of Parts 1 and 2 is recorded in seconds. The latency score for Part 2 was used for analysis.

Computational span—This task was used as an indicator of working memory requiring storage and simultaneous processing of information (Salthouse & Babcock, 1991). In the context of executive functioning, it represents the process of updating. Participants are asked to solve a series of arithmetic problems while holding the final digit from each problem in memory for later recall. The number of problems in a series increases from one to seven, with three trials at each series length. The highest span correctly recalled for two out of three trials was the measure used.

Reading span—This verbal task is similar to computational span in its requirement of updating. The reading span task requires participants to answer questions about simple

sentences that were orally presented while simultaneously trying to remember the final word of each sentence for later recall. The number of sentences in a passage increases from one to seven with three trials at each series length. The target measure was the highest span correctly recalled for two out of three trials.

Cognitive Reference Measures

We used indicators of five cognitive domains (i.e., perceptual speed, inductive reasoning, episodic memory, verbal fluency, and semantic memory) to evaluate the current cognitive status of each participant. The procedures have been established elsewhere (Dixon et al., 2007), and the reliability and validity of these measures are also discussed elsewhere (e.g., Hultsch et al., 1998).

Perceptual speed—Perceptual processing speed was assessed with the Wechsler Adult Intelligence Scale—Revised Digit Symbol Substitution (DSS) task (Wechsler, 1981). The DSS was selected as it represents speed, and its characteristics are well known in aging and other populations (e.g., MacDonald, Hultsch, & Dixon, 2003). The DSS is a relatively complex speed measure known to involve perceptual speed, motor speed, and working memory (Joy, Kaplan, & Fein, 2004). Participants are presented with a coding key pairing nine numbers (1 through 9) with nine different symbols. Printed under the coding key are rows of randomly ordered numbers with empty boxes below. Participants were given 90 s to transcribe as many symbols as possible into the empty boxes on the basis of the digit—symbol associations specified in the coding key. The number of correctly completed items represented the outcome measure.

Inductive reasoning—Inductive reasoning was assessed with the Letter Series test (Thurstone, 1962). Participants are presented with a string of letters forming a distinct pattern. The task requires inductively deciphering the pattern in the target string and providing the next letter in the string congruent with the pattern presented. The outcome measure used was the total number correct out of 20 patterns.

Episodic memory—The word recall task consists of immediate free recall of two lists of 30 English words selected from the total set of six lists (Dixon et al., 2004). Each list consists of 6 words from each of five taxonomic categories (e.g., birds, flowers) typed on a single page in unblocked order. Participants were given 2 min to study each list and 5 min to write their recall. The number of correctly recalled words averaged across the two lists was used as the measure.

Verbal fluency—Participants' verbal fluency was assessed with the Controlled Associations test from the Educational Testing Service (ETS) kit of factor-referenced cognitive tests (Ekstrom, French, Harman, & Dermen, 1976). The test requires the generation of as many synonyms as possible in response to a set of target words. Participants were given 6 min to complete the test, with the total number of correct synonyms representing the fluency score.

Vocabulary—A recognition vocabulary test represented semantic memory. The 54-item multiple-choice test was composed by concatenating three 18-item tests from the ETS kit of factor-referenced cognitive tests (Ekstrom et al., 1976). Participants were given 15 min, with the total number of correct items representing the vocabulary score.

Analyses

Alpha levels of $p < .05$ were specified as the threshold to indicate statistical significance. The CTT-2 latency and Stroop interference scores were reverse coded to be commensurate

with all other cognitive tasks (i.e., higher score = better performance). Descriptive data for all six EF tests are listed in Table 2. Correlations between the EF tests are presented in Table 3.

For the structural analyses we followed accepted conventions and our own previous research on this topic (de Frias et al., 2006). All structural models were run with the LISREL 8.80 program (Jöreskog & Sörbom, 2000) by using maximum likelihood estimation. Several indexes of model fit were considered in addition to the chi-square test, which is often significant in large samples. These indexes are the comparative fit index (CFI; Bentler, 1990), the goodness-of-fit index (GFI; Jöreskog & Sörbom, 2000), and the root-mean-square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger, 1990). Typically, RMSEA values at or below .05 are considered acceptable (although some authors accept values between .05 and .08; e.g., Browne & Cudeck, 1992). Generally, CFI and GFI values greater than .95 are indicative of good model fit (Hu & Bentler, 1999). The RMSEA, CFI, and GFI values range from 0 to 1. The chi-square statistic tests the closeness of fit between the sample covariance matrix (i.e., the data) and the fitted covariance matrix (i.e., the model). A nonsignificant p value for chi-square means that there is no reliable difference between the two matrices (i.e., the model is a good representation of the data) and therefore is one form of evidence toward accepting a model (especially in smaller samples). In larger samples (e.g., $N > 200$), the chi-square will likely be significant because of the greater sensitivity to detect smaller (and often trivial) differences between the sample covariance matrix and the fitted covariance matrix. As a result, exclusive reliance on the significance level of the chi-square test is not practiced. The CFI compares the predicted or hypothesized model with the (baseline) independence model (i.e., observed variables are uncorrelated). The GFI is an absolute index of fit that compares how much better a hypothesized model fits compared with no model. The RMSEA tests how well the model fits the population covariance matrix. Four sets of latent variable analyses (coordinated with the first four research goals) were conducted.

Research Goal 1: Initial structure of EFs—Confirmatory factor analyses were used to test the underlying structure of the six indicators of EF at W1 for each of the cognitive status groups separately. In one model, a three-factor structure of inhibition (Hayling overall scaled score, Stroop interference), shifting (Brixton overall scaled score, CTT-2 latency), and updating (Reading Span, Computational Span) was tested. In a second model, a single-EF factor was tested with all six indicators within each group. Factor scaling was established by fixing the factor loading of one item from the factor to a value of 1.0.

Research Goal 2: Group invariance of EF structure—The sequence of testing measurement invariance for the best-fitting structural model consisted of tests of configural and metric invariance (Horn & McArdle, 1992; Meredith, 1993). Based on the results of Research Goal 1, multigroup (cognitive status: CE, CN, CI) invariance was tested at W1. Configural invariance requires that the same variables load on the same number and pattern of factors across groups (Horn & McArdle, 1992). Metric invariance requires the equivalence of factor loadings between groups, and is employed here as a follow-up to confirmed configural invariance.

Research Goal 3: Longitudinal invariance of EF structure—Based on the results of the previous analyses, longitudinal (W1 to W2) invariance for each group was tested using the same methodology as in Research Goal 2 above.

Research Goal 4: Implications of stability in classification on structure and invariance of EFs—The previous analyses were based on group classifications at W1.

The present analyses began with the classification procedures being repeated for W2 data using the same methodology as in W1. Stability of group classification across the two waves was then evaluated. Within each group, those who were stable in their classifications were compared to those who changed classification. Structural and invariance models of EFs were tested between these stable–nonstable subgroups of the main cognitive status groups (CE, CN, and CI). Parenthetically, selection on the basis of extreme scores at W1 may be problematic when examining scores at follow-up, and especially with single- and low-reliability indicators. However, regression (and egression) to the mean is less problematic in the present situation, as participants are not selected on the basis of single extreme scores and the cognitive reference tasks have low-measurement error. Specifically, the retest correlations for the five classification tests are relatively high (Vocabulary, $r = .86$; Letter Series, $r = .81$; Digit Symbol Substitution, $r = .85$; Word Recall, $r = .79$; and Fluency, $r = .71$).

Research Goal 5: (Non)stability in classification and EF factor performance—

Given observed nonstability in classification across the two waves, EF performance differences were examined at both waves. Within each main cognitive status group (CE, CN, and CI), the subgroups (stable, nonstable) were compared in EF factor performance at each wave. A series of two (cognitive status) univariate and multivariate analyses of variances (M/ANOVAs) were conducted at W1 and W2 separately, in addition to 2 (cognitive status) $\times$ 2 (wave) repeated measures ANOVAs on the EF factor(s) (standardized to t -score units) as observed in previous analyses.

Results

Descriptive Univariate Analyses of EF Tasks

Prior to examining our main research questions, we conducted descriptive univariate analyses of the individual EF tasks (e.g., retest correlations) for the full sample. Results are presented in Table 4. There was overall 3-year improvement in performance on two tasks, Stroop: $t = -2.37$, $df = 340$, $p < .05$; Brixton: $t = -2.67$, $df = 334$, $p < .01$; and decline on three tasks, Hayling: $t = 3.41$, $df = 333$, $p < .001$; Reading Span: $t = 4.92$, $df = 343$, $p < .001$; Computational Span: $t = 3.82$, $df = 343$, $p < .001$. No significant change was observed for Color Trails. Three-year retest reliability correlations were moderate/high for all six tasks: Hayling ($r = .43$), Stroop ($r = .44$), Brixton ($r = .35$), Color Trails ($r = .65$), Reading Span ($r = .54$) and Computational Span ($r = .61$). As shown in Table 4, age correlated inversely and significantly with all executive functioning measures at both waves, indicating lower performance for older individuals. Years of education had a nonsignificant-to-low (positive) correlation (range = .01 to .22) with the executive measures.

Research Goal 1: Initial Structure of EFs

Confirmatory factor analyses were used to test the structure of the six indicators of EFs at W1. A multifactor model was tested against a single-factor model. A three-factor model of inhibition, shifting, and updating proved to be a good fit to the data for the CE group, $\chi^2(6, N = 77) = 5.55$, $p = .48$, RMSEA = .00, CFI = 0.98, GFI = 1.00. The interfactor correlations were moderate to high; however, the confidence intervals around the correlations were all nonsignificant, inhibition and updating, $r = .53$, $t = 1.06$, 95% confidence interval (CI) = $-.45, 1.52$; inhibition and shifting, $r = .89$, $t = .81$, 95% CI = $-1.28, 3.07$; updating and shifting, $r = .56$, $t = 1.07$, 95% CI = $-.47, 1.59$; likely a result of the low sample size ($n = 77$). The same three-factor model did not converge for the CI group, but did fit for the CN group, $\chi^2(6, N = 276) = 5.11$, $p = .53$, RMSEA = .00, CFI = 1.00, GFI = .99. However, for the latter, the correlation between pairs of factors (i.e., inhibition and shifting, $r = .97$; inhibition and updating, $r = .97$) was high (although $t < 2$) indicating that the factors were

not completely independent (although the shifting and updating correlation was acceptable, $r = .47$, $t > 2$). An alternative model that specified a single factor of EFs fit the data well for each group: CE, $\chi^2(9, N = 77) = 6.53$, $p = .69$, RMSEA = .00, CFI = 1.00, GFI = .97, PHI (“factor variance”) = 44%; CN, $\chi^2(9, N = 276) = 17.11$, $p = .05$, RMSEA = .057, CFI = 0.94, GFI = .98, PHI = 55%; and CI, $\chi^2(9, N = 120) = 5.63$, $p = .78$, RMSEA = .00, CFI = 1.00, GFI = .98, PHI = 60%. The standardized factor loadings for the single-factor model of EFs are presented in Table 5.

In summary, at W1 there was support for multidimensionality of EFs for CE adults, with the partial support for CN adults requiring more qualification. For the CI adults, a unitary model was a better fit. A single-factor model of EFs fit all three cognitive status groups. We used these results to guide the analyses in the next two research goals.

Research Goal 2: Group Invariance of EF Structure

We first tested invariance of the three-factor solution, comparing the CE and CN groups. The three-factor model was not invariant from the CE to CN groups (i.e., it did not converge). Next, because the one-factor model fit (moderately for all groups) we estimated configural invariance of the one-factor model for the CE, CN, and CI groups in a series of two-way comparisons at W1. For the CE and CN group comparisons, the model fit the data well, $\chi^2(18, N = 353) = 23.64$, $p = .17$, RMSEA = .04, CFI = .96, GFI = .97; indicating that the pattern of loadings and the number of factors is similar for these groups. Metric invariance of factor loadings was tested by constraining the factor loadings between CE and CN adults. This model also fit the data equally well, $\chi^2(23, N = 353) = 31.41$, $p = .11$, RMSEA = .05, CFI = .95, GFI = .95; and the chi-square difference test that compares a more constrained model with a less constrained one did not compromise model fit, $\Delta\chi^2(5) = 7.77$, $p > .05$. The magnitude of the relations between the observed variables and the factor is similar for both groups.

For the CN and CI group comparisons, configural invariance of the one-factor model fit the data well, $\chi^2(18, N = 396) = 22.73$, $p = .20$, RMSEA = .04, CFI = .98, GFI = .98. Metric invariance of factor loadings was tested by constraining the factor loadings between CN and CI adults. This model also fit the data equally well, $\chi^2(23, N = 396) = 31.81$, $p = .10$, RMSEA = .04, CFI = .96, GFI = .97; $\Delta\chi^2(5) = 9.08$, $p > .05$.

For the CE and CI group comparisons, configural invariance of the one-factor model fit the data well, $\chi^2(18, N = 197) = 12.16$, $p = .84$, RMSEA = .00, CFI = 1.00, GFI = .97. Metric invariance of factor loadings was tested by constraining the factor loadings between CE and CI adults. This model did not fit the data, $\chi^2(23, N = 197) = 25.80$, $p = .31$, RMSEA = .04, CFI = .97, GFI = .94; $\Delta\chi^2(5) = 13.64$, $p < .05$; indicating that the magnitude of the relations between the observed variables and the single factor are different for CE and CI adults. The goodness of fit statistics are presented in Table 6.

In summary, the three-factor model was not invariant between the CE and CN groups, indicating its uniqueness to the CE group. The single-factor model of EFs however was invariant (configural and metric) for the CE–CN and CN–CI comparisons, but not (i.e., metric) for the extreme Groups CE–CI comparison. These results confirm (a) the relative uniqueness of the three-factor model to the CE group and (b) the generalizability of one-factor model, as qualified by a lack of invariance between the cognitive status extremes (i.e., CE and CI).

Research Goal 3: Longitudinal Invariance of EF Structure

Based on the accumulating evidence for a three-factor model for the CE group, we first estimated configural invariance across the 3-year (two-wave) period for this group. The

three-factor model fit the data well, $\chi^2(33, N = 62) = 28.43, p = .69$, RMSEA = .00, CFI = .99, GFI = .93. Metric invariance of factor loadings across the 3-year period was tested by constraining the factor loadings between W1 and W2. This model also fit the data, $\chi^2(39, N = 62) = 33.05, p = .74$, RMSEA = .00, CFI = 1.00, GFI = .92; $\Delta\chi^2(6) = 4.62, p > .05$; indicating that the magnitude of the relations between the variables and the three factors is similar at W1 and W2 for the CE group. We examined longitudinal invariance of the three-factor solution for the CN group. The model fit well at the configural level, $\chi^2(33, N = 190) = 38.28, p = .24$, RMSEA = .03, CFI = .99, GFI = .97; and at the metric level, $\chi^2(39, N = 190) = 48.09, p = .15$, RMSEA = .04, CFI = .99, GFI = .96; $\Delta\chi^2(6) = 9.81, p > .05$.

We examined longitudinal invariance for the single-factor model separately for each group. For the CE group, configural invariance of the single-factor model did not fit the data well, $\chi^2(47, N = 62) = 60.71, p = .09$, RMSEA = .07, CFI = .94, GFI = .86. Comparing across the two sets of analyses for the CE group, the three-factor model fit the data better than did the single-factor model. In contrast, for the CN group, configural invariance of the single-factor model fit the data well, $\chi^2(47, N = 190) = 67.88, p = .02$, RMSEA = .05, CFI = .97, GFI = .94; indicating that the structure of EF is similar at W1 and W2. However, metric invariance of the same factor was not supported, $(\chi^2(52, N = 190) = 84.23, p = .00, RMSEA = .057, CFI = .95, GFI = .93; \Delta\chi^2(5) = 16.35, p < .05$; indicating that the size of the relations between the variables and the single factor is different at W1 and W2. For the CI group, configural invariance of the single-factor model fit the data well, $\chi^2(47, N = 71) = 55.50, p = .18$, RMSEA = .05, CFI = .95, GFI = .88; indicating that the structure of EF is similar at both waves. Metric invariance of the single factor was also supported by the data, $\chi^2(52, N = 71) = 65.11, p = .10$, RMSEA = .06, CFI = .94, GFI = .87; $\Delta\chi^2(5) = 9.61, p > .05$; indicating that the size of the relations between the variables and the single factor is similar at both waves. The goodness of fit statistics are presented in Table 6.

In summary, the three-factor model of EFs was longitudinally invariant (metric and configural) over the 3-year period for the CE and CN groups. The single-factor model was robustly invariant (configural and metric) only for the CI group, as mixed results were obtained for the CN group (i.e., no metric invariance). Contrary results were observed for the CE group.

Research Goal 4: Implications of Stability in Classification on Structure and Invariance of EFs

New group classifications were conducted using all individuals who returned for W2, and stability was examined. Based on the W1 classification, the cognitive status groups were moderately stable over the 3-year period: CE (60.8%; 31 of 51 returners), CN (76.5%; 137 of 179 returners), and CI (72.2%; 57 of 79 returners). The differences between the proportions were not statistically significant ($p > .05$). Of the roughly half of each group who returned but changed status at W2, nearly all were one-step changes: (a) CE (100%; 20 of 20 nonstable adults changed “down” to CN), (b) CN (76.2%; 32 of 42 changed “down” to CI), and (c) CI (95.5%; 21 of 22 changed “up” to CN). For the present and next series of analyses we tested whether the stable group members (i.e., CE-to-CE, CN-to-CN, CI-to-CI, from W1 to W2, respectively) were different in structure (Research Goal 4) and performance (Research Goal 5) in EF.

The stable CE-to-CE group had too low a sample size ($n = 31$) to estimate reliable structural models. Structural models were estimated for CN-to-CN and CI-to-CI adults who were stable in their group status classification over the 3-year period. Under the assumption that the stable CN group would be functionally closer to CE adults (i.e., near elite cognitively) than to CI adults, we first tested the stability of the three-factor model of EFs, the one that fit best for the CE group, but somewhat inconsistently for the overall CN group. This model

was a good fit for this more select (i.e., 3-year stable) group of initially CN participants. The model fit the data well, $\chi^2(6, N = 130) = 3.64, p = .73$, RMSEA = .00, CFI = 1.00, GFI = .99; and the interfactor correlations were appropriately moderate (shifting and inhibition, $r = .77$; inhibition and updating, $r = .79$; shifting and updating, $r = .31$). Longitudinal configural invariance of the three-factor model fit the data well, $\chi^2(33, N = 130) = 44.80, p = .08$, RMSEA = .05, CFI = .98, GFI = .95; as did metric invariance, $\chi^2(39, N = 130) = 51.70, p = .08$, RMSEA = .05, CFI = .97, GFI = .94; $\Delta\chi^2(6) = 6.90, p > .05$. By comparison, the three-factor model for the stable CI-to-CI group did not converge.

The one-factor model of EFs at W1 was re-estimated in the stable CN-to-CN group, and it was still a good fit to the data, $\chi^2(9, N = 130) = 8.82, p = .45$, RMSEA = .00, CFI = 1.00, GFI = .98. Configural invariance of the single-factor model across the 3-year period did not fit the data, $\chi^2(47, N = 130) = 61.58, p = .08$, RMSEA = .08, CFI = .91, GFI = .82. For the lowest cognitive classification, the stable CI-to-CI group, the single-factor model was not a good fit to the data, $\chi^2(9, N = 48) = 14.04, p = .12$, RMSEA = .11, CFI = .77, GFI = .91. Those who continued as impaired over the 3 years display a less consistent pattern of executive functioning.

In summary, the near-elite group (i.e., stable CN-to-CN individuals) displayed a three-factor pattern of EF structure that was quite similar to that observed earlier for their W1 CE counterparts and different from that observed earlier for the overall CN group. The least cognitively elite group (CI-to-CI) produced no evidence of longitudinal structural invariance in EF. There were too few stable CE individuals ($n = 31$) to test a reliable model, but the earlier W1 analyses indicated a three-factor solution similar to that of the CN-to-CN group at W2.

Research Goal 5: (Non)Stability in Classification and EF Factor Performance

We considered the possibility that the stable subgroups (e.g., CE-to-CE) may differ in EF performance from the corresponding nonstable subgroups (e.g., CE-to-CN). Accordingly, we conducted a series of analyses using the observed factor structures (three-factor solution for the CE groups; one-factor solution for all groups) to test whether the corresponding stable and nonstable subgroups performed differently in level (as measured in factor summary scores). Our prediction is that the stable CE and CN groups would perform better than their counterparts (who dropped a level) and that, correspondingly, the stable CI group would perform worse than its counterpart (who increased a level). We elected to test this prediction with one-way (multivariate) ANOVAs at each wave separately. Factor scores were created by using unit weighting (summing across standardized variables). We first conducted a series of 2 (cognitive status) $\times$ 2 (wave) repeated-measures ANOVAs. As significant group effects, but no significant wave effects or group by wave interactions were observed, we performed one-way ANOVAs at each wave.

We began with the CE group, testing performance on the three factors with a series of 2 (stability groups: CE-to-CE, CE-to-CN) multivariate ANOVAs conducted at W1 and W2 separately. The results showed a significant group effect at W1 for updating, $F(1, 49) = 19.20, p < .001, \eta^2 = .28$. As shown in Figure 1 (panel A), the stable CE-to-CE group ($M = 61.79, SD = 9.80$) performed better than the nonstable CE-to-CN group ($M = 50.50, SD = 7.51$). The effect of group was significant for all three EF domains at W2: inhibition, $F(1, 49) = 4.89, p < .05, \eta^2 = .09$; shifting, $F(1, 49) = 5.87, p < .05, \eta^2 = .11$; and updating, $F(1, 49) = 20.14, p < .001, \eta^2 = .29$. In all cases, the stable CE-to-CE group (inhibition: $M = 53.93, SD = 4.98$; shifting: $M = 55.10, SD = 5.24$; and updating: $M = 59.42, SD = 8.42$) performed better than the nonstable CE-to-CN group (inhibition: $M = 50.61, SD = 5.62$; shifting: $M = 50.54, SD = 8.23$; and updating: $M = 49.08, SD = 7.38$).

Next, we tested performance on the single EF for the stable versus nonstable subgroups within each overall group. The first comparison was between the stable CE-to-CE group ($n = 31$) and the nonstable CE-to-CN group ($n = 19$). There was a significant group difference at W1, $F(1, 48) = 13.88$, $p < .001$, $\eta^2 = .22$; and at W2, $F(1, 48) = 17.01$, $p < .001$, $\eta^2 = .26$. The stable CE-to-CE group (W1: $M = 56.43$, $SD = 5.22$; W2: $M = 56.15$, $SD = 4.48$) performed better than the CE-to-CN group (W1: $M = 51.09$, $SD = 4.38$; W2: $M = 50.31$, $SD = 5.43$) at both waves (see Figure 2, panel A). The second comparison was between the stable CN-to-CN group ($n = 132$) and the nonstable CN-to-CI group ($n = 32$). There was a significant group difference at W1, $F(1, 162) = 5.72$, $p < .05$, $\eta^2 = .03$; and at W2, $F(1, 163) = 5.36$, $p < .05$, $\eta^2 = .03$. The stable group (W1: $M = 50.22$, $SD = 5.07$; W2: $M = 49.89$, $SD = 5.35$) performed better than the CN-to-CI group (W1: $M = 47.83$, $SD = 5.09$; W2: $M = 47.36$, $SD = 6.05$) at both waves (see Figure 2, panel B). The third comparison was between the stable CI-to-CI group ($n = 50$) and the nonstable CI-to-CN group ($n = 20$). There was a significant group difference at W1, $F(1, 68) = 3.96$, $p < .05$, $\eta^2 = .06$; and W2, $F(1, 73) = 7.74$, $p < .01$, $\eta^2 = .10$. As expected, the nonstable CI-to-CN group (W1: $M = 50.13$, $SD = 5.17$; W2: $M = 50.38$, $SD = 3.76$) performed better than the stable CI-to-CI group (W1: $M = 47.38$, $SD = 5.24$; W2: $M = 45.52$, $SD = 7.63$) at both waves (see Figure 2, panel C).

In summary, we accounted for the (non)stability of cognitive classification by comparing performance for those who remained in group (a positive sign for CE-to-CE and CN-to-CN subgroups, but a negative sign for the CI-to-CI subgroup). The results followed the predictions, in that the expected higher level cognitive subgroups performed better than their corresponding subgroups on the single-EF factor. The stable CE-to-CE group, performed better than the nonstable CE-to-CN group across all three executive factors (i.e., W1: updating; and W2: all three domains).

Discussion

The overall purpose of this research was to examine the structure, stability, and performance of EFs in three select groups of older adults. From an initial sample of over 500 older participants, we developed three nonoverlapping but conceptually continuous cognitive status groups. Research in structure and performance of EF has been conducted with presumably neurologically intact younger adults or cognitively typical older adults (e.g., de Frias et al., 2006; Ettenhofer et al., 2006; Miyake et al., 2000). We empirically established a cognitively normal older group (i.e., CN), and compared their performance characteristics to those of two unique older groups at both the cognitively healthier (i.e., CE) and cognitively less healthy (i.e., CI) borders. Structural studies of EF have not previously targeted the cognitively elite or mildly impaired groups, although some studies of EF performance have produced valuable information for representatives of the latter populations (e.g., Rapp & Reischies, 2005; Zhang et al., 2007). We used general perspectives on EF and previous structural studies to develop a set of five research goals and several associated expectations. The general guiding principle was that a more differentiated structure of EF would be observed for the more cognitively intact or elite group (i.e., CE), consistent with the notion that their executive functioning may resemble that of younger adults, and differ from that of CN and CI older adults. Because each of the five research goals approaches this guiding principle with coordinated data, we discuss the overall supportive results for each segment sequentially.

Research Goals 1 and 2 were pursued to establish the initial factor structures of our EF battery and to test the measurement invariance of those structures across each of the groups. Using the W1 sample, we observed that the three-factor structure (inhibition, shifting, updating) was indeed confirmed for the CE group and clearly rejected for the CI group, with the intermediate CN group producing qualified confirmatory evidence. The one-factor

model of EF fit the data for each group. Although both the one-factor and three-factor models fit the CE group, it is generally the case that simpler statistical models are easier to fit than complex models (i.e., those estimating more parameters). One indirect inference is that empirical support for the more complex model is a testament to executive function differentiation in healthy older brains. Of note for the three-factor result is that the interfactor correlations were moderate to high, a fact that is reflected in the complementary one-factor model. Therefore, there is support for diversity of EF in the CE group, although this is qualified by shared variances in factors. A post hoc follow-up second-order model with the three first-order factors represented by a single higher order factor (called EF) fit the data well, $\chi^2(6, N = 77) = 5.55, p = .48, RMSEA = .00, CFI = 1.00, GFI = .98$. (The correlations between the second-order factor and first-order factors were moderate to high ($r = .58$ to $.97$), but not significant ($t < 2$). Given that our CE group is selected within the context of a study of initially normal aging, it is conceivable that a group selected more specifically on the basis of extreme cognitively elite aging could show even stronger evidence of the complex structure of executive functioning. Overall, the data support dedifferentiation of executive functions (i.e., three factors for CE, two factors for CN, and one factor for CI).

When testing group invariance, the models revealed (a) a lack of invariance for the three-factor solution (comparing CE and CN), (b) evidence for invariance for the one-factor solution between the CE–CN and CN–CI comparisons, but (c) a lack of invariance for the (extreme-group) CE–CI comparison. Taken together, this evidence suggests that the construct of EF is, as expected, more differentiated in the CE group, and that the multidimensional structure is relatively unique to this group among older adults. Underlying this select multidimensionality, however, is the fact that the EF tasks hold together in a one-factor model that fits for all three groups. Thus, large samples of older adults (such as those we reported earlier by de Frias et al., 2006), may produce evidence of homogeneity of EF structure, especially if they are dominated by relatively normal (i.e., typically cognitively declining) and impaired (i.e., declining more rapidly although not necessarily into a neurodegenerative disease) adults—as are many of the major longitudinal studies. Future longitudinal research could examine the extent to which the CE adults may be resistant to the dedifferentiation that could be occurring with aging for the CN and CI adults. At present, our results are consistent with the interpretation that the cognitively elite older adults are not only broadly intact but deeply intact. This is indicated by their differentiated structure of executive functioning, a pattern that appears to exist in healthy young adults (e.g., Miyake et al., 2000).

Whereas the first two research goals addressed the invariance of EF structure across different groups on a single occasion, the third research goal pursued the novel question (in EF and aging research) of whether the observed structures would be temporally invariant within each group. The VLS provided a unique opportunity for a strong test of this temporal invariance hypothesis, as the between-wave interval is three years. This period for older adults would be sufficient for unreliability or instability to begin appearing. Because of the clear evidence for a three-factor structure for the CE group at W1, and the marginal evidence for the CN group, we tested temporal invariance of the multidimensional structure for these two groups across W1 and W2. The models fit well for both groups, at both the configural and metric levels. Complementing these results, the one-factor longitudinal invariance model was a poor fit for the CE group but a good fit for the CI group. The results for the CN group were again somewhat intermediate to those obtained for the other two groups, in that the three-factor model appeared to fit best but the one-factor model was also configurally invariant (and it was not for the CE group). One limitation is that the lack of invariance could be due to the smaller sample sizes for the longitudinal samples (cf. cross-sectional W1 samples).

Why might the CN group produce evidence supportive of longitudinal invariance for the three-factor model? According to our theory, we had expected this result for the CE group (and found it), but our previous work with a relatively normal older adult sample (de Frias et al., 2006) led us to expect that the CN group would be more typified by a unidimensional than a multidimensional structure of EF. The results of our first three research goals were moderately consistent with this latter expectation, but the remaining ambiguity about the EF structure for CN adults led us to pursue the fourth research goal. Specifically, for the fourth goal we examined the possibility that the nonstability of cognitive status across time could be associated with our results. When multiple measures are given, it is not uncommon for even neurologically normal individuals to show some abnormal test performances (e.g., Heaton, Miller, Taylor, & Grant, 2004; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008). These isolated low scores may reflect a variety of sources, including chance factors, statistical artifacts of the procedure used to define abnormality, motivational defects as well as indications of mild cognitive impairment. In particular, it is possible that some CN individuals may be robustly CN whereas others may be transitioning to a pre-CI status. Therefore, we reclassified the older adults who returned for W2, and then computed stability information. The stability of classification was relatively high (cf. with those observed in previous research with mild cognitive impairment by Garrett, MacDonald, & Dixon, 2006; Palmer, Wang, Bäckman, Winblad, & Fratiglioni, 2002). Although the stable CE-to-CE group was too small to support structural modeling, the stable CN-to-CN group presented a unique opportunity to test a group that is arguably closest (empirically and conceptually) to their slightly more intact counterparts. Indeed, the three-factor structure was confirmed for this stable and relatively cognitively maintained group. One additional notable finding was that the stable CI-to-CI group produced a poor fit for even the one-factor model of EF. This group is arguably the least cognitively intact of the various subgroups, as no further decline is defined or measured in this study. Although no structural EF research has previously been conducted with mildly cognitively impaired older adults, the present results suggest that (a) the construct of EF may not be valid once the process of accelerated impairment has begun, (b) the results may reflect executive dysfunction perhaps due to degeneration in the neural substrates subserving EF functioning (e.g., Stuss & Benson, 1984), (c) researchers and clinicians using single or few indicators of EF with such populations may not produce valid measurements of executive functioning or dysfunction, and (d) future large-sample and multivariate research on such clinical (preferably multiple homogenous patient groups) and preclinical populations is recommended.

For our fifth and final research goal, we switched from the structural questions to a restricted set of level-of-EF performance questions. Specifically, we tested whether those independently classified groups that were subsequently subdivided by 3-year stability status would perform on the appropriate EF factors in a predictable fashion. The overall prediction would be that the combination of higher classification and more stability would produce subgroups that would perform better on the EF factors. To illustrate, we observed that the CE-to-CE group generally performed better on the three EF factors than did the CE-to-CN group. As the three-factor model applied primarily to the better-performing groups, we tested the EF performance on the one-factor model across all comparisons. All predictions were confirmed: (a) the CE-to-CE group performed better on mean EF single-factor scores than did the CE-to-CN group, (b) the CN-to-CN group performed better than did the CN-to-CI group, and (c) the CI-to-CN group performed better than the stable CI-to-CI group. Thus, our derived factor score representations of EF performance tracked the independent classifications of cognitive status and the objective information about the 3-year stability of these classifications.

Overall, the results are notably consistent with our expectations and present a theoretically coherent picture of executive functioning in special populations of older adults. Despite the

robustness of the results, several limitations to the present study should be noted. First, a set of methodological limitations is present in most EF research, especially the selection of EF measures (from the cognitively pure to the neuropsychological tests). We used a combination of such indicators and performed our major analyses at the latent variable level, thus mitigating some of the appropriate psychometric concerns. In addition, the methodological strengths of multiple indicators, large samples, and longitudinal data should also be noted. Noteworthy, a different set of measures (i.e., from the same construct or of different psychometric qualities) may have produced a different number of factors. Second, although our groups are independently classified following published procedures, they are not necessarily externally valid. Specifically, we use the term cognitively elite to refer to our better performing individuals, but we have no external criteria by which we could establish that they are indeed elite at a population level. Similarly, we use the term cognitively impaired to refer to our lowest cognitive status group, but this does not imply that they should be considered as diagnosed or confirmed cases of “mild cognitive impairment” or linked with a particular neurodegenerative disease (see Dixon et al., 2007). Instead, we use these terms to refer to relatively better and relatively poorer performers within this sample (rather than misleadingly treating them as a homogenous group), relying on the precision of our established procedures to support the distinction. Future research on diagnosed special populations (such as mild cognitive impairment or Alzheimer’s disease) is encouraged. The classification of cognitive status was derived from performance on five basic cognitive domains, rather than representing cognitive ability along a continuum. Our multivariate approach allows for follow-up analyses on differential sensitivities to aging and impairment that has clinical and theoretical importance. Third, selection on extreme scores and apparent regression to the mean at W2 may have some influence on the mean-level results from Goals 4 and 5. However, we should note that (a) multiple indicators were used not single low or high scores, (b) retest correlations for the five classification tests were relatively high, suggesting low measurement error, and (c) there was some movement by a minority of the CE and CI groups in the direction consistent with regression/egression but the possibility of movement in the other direction was not defined and not coded. A third wave would help resolve this question, but such information will not be available in the VLS for a couple of years. Fourth, given our classification procedures (in which individuals are classified according to relative performance on one of five cognitive reference tests), it is not possible to evaluate the magnitudes of the shifts involved in the categorical status changes. If there were a single dimension (of, say, cognitive competence), then measuring the magnitudes of the shifts could be possible. Fifth, the present research has no direct neurobiological indicators. Imaging data, especially, might help explain the mechanisms by which behavioral evidence of EF structure or EF performance are affected by normal or impaired aging (e.g., extent of frontal lobe involvement; Stuss & Benson, 1984). Equally interesting for future research would be the neurobiological correlates of the complex structure and high levels of performance produced by the CE older adults. Unknown to date is whether these phenomena would be produced by neurobiological phenomena equivalent to that of younger adults or by compensatory neurobiological mechanisms (Buckner, 2004; Dixon et al., 2008; Reuter-Lorenz & Cappell, 2008). Sixth, future studies with additional longitudinal waves may wish to examine dynamic relations between changes in cognitive status (along a continuum) and changes in executive functioning (e.g., differentially by dimension).

In sum, we found that with respect to EF structure and EF performance, older adult groups can be usefully distinguished in terms of their relative cognitive intactness. Cognitively elite older adults represent those who may be aging in a selectively healthy or successful manner (Dixon, 2009). Arguably, they may have preserved (or otherwise attained) a structure and level of executive functioning similar to that of younger adults (or perhaps to themselves as younger adults). Depending on their relative cognitive status—for example, whether they are tending in the short term toward maintenance or toward impairment—normal aging adults

may occupy an EF status intermediate to that of the CE and CI adults. Finally, CI adults seem to show a single dimension of EF, at least as measured by the present tests. Moreover, stable CI adults appear to experience dissolution of EF structural integrity, presaging perhaps a more accelerated or neuropathogenic decline, possibly to the point of executive dysfunction with everyday consequences. Further research on actual changes and neurocognitive correlates of EF in each of these special populations is recommended.

Acknowledgments

This research was supported by a grant from the National Institutes of Health (National Institute on Aging) to Roger A. Dixon (R37 AG008235). Roger Dixon was also supported by the Canada Research Chairs program. Further information about the Victoria Longitudinal Study may be accessed via: <http://www.ualberta.ca/~vlslab/index.html>

References

- Amieva H, Phillips LH, Della Sala S, Henry JD. Inhibitory functioning in Alzheimer's disease. *Brain*. 2004; 127:949–964. [PubMed: 14645147]
- Bentler PM. Fit indices, LaGrange multipliers, constraint changes, and incomplete data in structural models. *Multivariate Behavioral Research*. 1990; 25:163–172.
- Bielak AAM, Mansueti L, Strauss E, Dixon RA. Performance on the Hayling and Brixton tests in older adults: Norms and correlates. *Archives of Clinical Neuropsychology*. 2006; 21:141–149. [PubMed: 16242905]
- Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociological Methods & Research*. 1992; 21:230–258.
- Buckner RL. Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*. 2004; 44:195–208. [PubMed: 15450170]
- Burgess PW, Alderman N, Forbes C, Costello A, Coates LM-A, Dawson DR, et al. The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*. 2006; 12:194–209. [PubMed: 16573854]
- Burgess, PW.; Shallice, T. *The Hayling and Brixton Tests*. Thurston, England: Thames Valley Test Company; 1997.
- de Frias CM, Dixon RA, Strauss E. Structure of four executive functioning tests in healthy older adults. *Neuropsychology*. 2006; 20:206–214. [PubMed: 16594781]
- D'Elia, LF.; Satz, P.; Uchiyama, CL.; White, T. *Color Trails Test*. Odessa, FL: PAR; 1996.
- Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*. 2003; 9:936–946. [PubMed: 14632252]
- Dixon, RA. An epidemiological approach to cognitive health in aging. In: Bäckman, L.; Nyberg, L., editors. *Memory, aging, and the brain*. London: Psychology Press; 2009. p. 144–166.
- Dixon RA, de Frias CM. The Victoria Longitudinal Study: From characterizing cognitive aging to illustrating patterns and predictors of changes in memory compensation. *Aging, Neuropsychology, and Cognition*. 2004; 11:346–376.
- Dixon, RA.; Garrett, DD.; Bäckman, L. Principles of compensation in cognitive neuroscience and neurorehabilitation. In: Stuss, DT.; Winocur, G.; Robertson, IH., editors. *Cognitive neurorehabilitation*. 2. Cambridge, England: Cambridge University Press; 2008. p. 22–38.
- Dixon RA, Garrett DD, Lentz TL, MacDonald SWS, Strauss E, Hultsch DF. Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*. 2007; 21:381–399. [PubMed: 17484601]
- Dixon RA, Wahlin Å, Maitland SB, Hultsch DF, Hertzog C, Bäckman L. Episodic memory change in late adulthood: Generalizability across samples and performance indices. *Memory & Cognition*. 2004; 32:768–778.

- Ekstrom, RB.; French, JW.; Harman, HH.; Dermen, D. Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service; 1976.
- Ettenhofer ML, Hambrick DZ, Abeles N. Reliability and stability of executive functioning in older adults. *Neuropsychology*. 2006; 20:607–613. [PubMed: 16938023]
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Friedman NP, Miyake A. The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*. 2004; 133:101–135. [PubMed: 14979754]
- Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK. Not all executive functions are related to intelligence. *Psychological Science*. 2006; 17:172–179. [PubMed: 16466426]
- Garrett, DD.; MacDonald, SWS.; Dixon, RA. Characteristics of the cognitive battery affect prevalence and stability of subclinical cognitive impairment; Paper presented at the Cognitive Aging Conference; Atlanta, GA. 2006 Apr.
- Gunning-Dixon FM, Raz N. The cognitive correlates of white matter abnormalities in normal aging: A quantitative review. *Neuropsychology*. 2000; 14:224–232. [PubMed: 10791862]
- Heaton, RK.; Miller, SW.; Taylor, MJ.; Grant, I. Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz, FL: Psychological Assessment Resources; 2004.
- Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*. 1992; 18:117–144. [PubMed: 1459160]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structural analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
- Hull R, Martin RC, Beier ME, Lane D, Hamilton AC. Executive function in older adults: A structural equation modeling approach. *Neuropsychology*. 2008; 22:508–522. [PubMed: 18590362]
- Hultsch, DF.; Hertzog, C.; Dixon, RA.; Small, BJ. Memory change in the aged. New York: Cambridge University Press; 1998.
- Jöreskog, KG.; Sörbom, D. LISREL 8 user's reference guide. Chicago: Scientific Software; 2000.
- Joy S, Kaplan E, Fein D. Speed and memory in the WAIS–III Digit Symbol-Coding subtest across the adult lifespan. *Archives of Clinical Neuropsychology*. 2004; 19:759–767. [PubMed: 15288329]
- Lezak, MD.; Howieson, DB.; Loring, DW. Neuropsychological assessment. 4. New York: Oxford University Press; 2004.
- Li S-C, Lindenberger U, Sikstrom S. Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*. 2001; 5:479–486. [PubMed: 11684480]
- Lowe C, Rabbitt P. Test–retest reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia*. 1998; 36:915–923. [PubMed: 9740364]
- Luszcz, MA.; Lane, AP. Executive function in cognitive, neuropsychological, and clinical aging. In: Hofer, SM.; Alwin, DF., editors. *The handbook of cognitive aging*. Thousand Oaks, CA: SAGE; 2008. p. 193–206.
- MacDonald SWS, Hultsch DF, Dixon RA. Age-related slowing of digit symbol substitution revisited: What do longitudinal age changes reflect? *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2003; 58:P187–P194.
- Marczewski P, Van der Linden M, Laroi F. Further investigation of the supervisory attentional system in schizophrenia: Planning, inhibition, and rule abstraction. *Cognitive Neuropsychiatry*. 2001; 6:175–192.
- Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*. 1993; 58:525–543.
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howeter A, Wager T. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*. 2000; 41:49–100. [PubMed: 10945922]

- Moritz S, Birkner C, Kloss M, Jahn H, Hand I, Haasen C, et al. Executive functioning in obsessive-compulsive disorder, unipolar depression, and schizophrenia. *Archives of Clinical Neuropsychology*. 2002; 17:477–483. [PubMed: 14592001]
- Palmer K, Wang H-X, Bäckman L, Winblad B, Fratiglioni L. Differential evolution of cognitive impairment in nondemented older persons: Results from the Kungsholmen project. *American Journal of Psychiatry*. 2002; 159:436–442. [PubMed: 11870008]
- Psychological Assessment Resources. Wisconsin Card Sorting Test: Scoring Program (version 3.0). Odessa, FL: 1990.
- Rabbitt P, Lowe C, Shilling V. Frontal tests and models for cognitive aging. *European Journal of Cognitive Psychology*. 2001; 13:5–28.
- Rapp MA, Reischies FM. Attention and executive control predict Alzheimer disease in later life: Results from the Berlin Aging Study (BASE). *American Journal of Geriatric Psychiatry*. 2005; 13:134–141. [PubMed: 15703322]
- Raz N, Gunning FM, Head D, Dupuis JH, McQuain J, Briggs SD, et al. Selective aging of the human cerebral cortex observed in vivo: Differential vulnerability of the prefrontal gray matter. *Cerebral Cortex*. 1997; 7:268–282. [PubMed: 9143446]
- Ready RE, Ott BR, Grace J, Cahn-Weiner DA. Apathy and executive dysfunction in mild cognitive impairment and Alzheimer disease. *American Journal of Geriatric Psychiatry*. 2003; 11:222–228. [PubMed: 12611752]
- Reitan, RM.; Wolfson, D. The Halstead-Reitan Neuropsychological Test Battery. Tucson, AZ: Neuropsychology Press; 1985.
- Regard, M. Unpublished doctoral dissertation. University of Victoria; Victoria, British Columbia, Canada: 1981. Cognitive rigidity and flexibility: A neuropsychological study.
- Reuter-Lorenz PA, Cappell K. Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Sciences*. 2008; 17:177–182.
- Ritchie K, Artero S, Touchon J. Classification criteria for mild cognitive impairment: A population-based validation study. *Neurology*. 2001; 56:37–42. [PubMed: 11148233]
- Royall DR, Lauterbach EC, Cummings JL, Reeve A, Rummans TA, Kaufer DI, et al. Executive control function: A review of its promise and challenges for clinical research. *Journal of Neuropsychiatry and Clinical Neurosciences*. 2002; 14:377–405. [PubMed: 12426407]
- Salthouse TA, Atkinson TM, Berish DE. Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*. 2003; 132:566–594. [PubMed: 14640849]
- Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. *Developmental Psychology*. 1991; 27:763–776.
- Schretlen DT, Testa SM, Winicki JM, Pearlson GD, Gordon B. Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society*. 2008; 14:436–445. [PubMed: 18419842]
- Sliwinski MJ, Hofer SM, Hall C. Correlated and coupled cognitive change in older adults with and without preclinical dementia. *Psychology and Aging*. 2003; 18:672–683. [PubMed: 14692856]
- Smitherman TA, Huerkamp JK, Miller BI, Houle TT, O’Jile JR. The relation of depression and anxiety to measures of executive functioning in a mixed psychiatric sample. *Archives of Clinical Neuropsychology*. 2007; 22:647–654. [PubMed: 17521865]
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
- Stern, Y., editor. *Cognitive Reserve*. New York: Psychology Press; 2006.
- Strauss, E.; Sherman, EM.; Spreen, O. A compendium of neuropsychological tests: Administration, norms, and commentary. 3. New York: Oxford University Press; 2006.
- Stuss DT, Benson DF. Neuropsychological studies of the frontal lobes. *Psychological Bulletin*. 1984; 95:3–28. [PubMed: 6544432]
- Stuss DT, Shallice T, Alexander MP, Picton TW. A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences*. 1994; 769:191–211. [PubMed: 8595026]

- Thurstone, TG. Primary mental abilities: Grades 9 –12, 1962 revision. Chicago: Science Research Associates; 1962.
- Tuokko, H.; Hultsch, DF., editors. Mild cognitive impairment: International perspectives. New York: Taylor & Francis; 2006.
- Wechsler, D. WAIS–R manual. New York: Psychological Corporation; 1981.
- Zhang Y, Han B, Verhaeghen P, Nilsson L-G. Executive functioning in older adults with mild cognitive impairment: MCI has effects on planning but not on inhibition. *Aging, Neuropsychology, and Cognition*. 2007; 14:557–570.

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1.

The stable CE-to-CE group performed better than the CE-CN group at Wave 1 (W1; updating) and Wave 2 (W2; inhibition, shifting, and updating). (All participants were CE at W1.) CE = cognitively elite; CN = cognitively normal.

Figure 2.

Panels A and B: The higher functioning stable groups (CE-to-CE, CN-to-CN) performed better on the single-EF factor than their counterparts who dropped in status. C: The stable CI-to-CI group performed worse than the group who changed status (CI-to-CN group). CE = cognitively elite; CN = cognitively normal; CI = cognitively impaired.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Background Information for Each Cognitive Status Group at W1 and W2

Variable	CE			CN			CI		
	M	SD	Range	M	SD	Range	M	SD	Range
Age									
W1	66.05	7.83	55 to 86	68.45	8.65	54 to 88	68.80	8.55	53 to 90
W2	70.06	7.65	59 to 89	72.62	8.62	58 to 92	72.55	7.86	57 to 94
Education									
W1	15.81	3.32	9 to 22	15.29	2.74	9 to 23	14.67	3.25	3 to 24
W2	15.68	3.12	10 to 23	15.74	2.76	8 to 22	15.06	2.95	5 to 23
MMSE									
W1	29.10	1.27	24 to 30	28.73	1.20	24 to 30	28.46	1.30	24 to 30
W2	29.29	0.87	27 to 30	28.48	1.26	24 to 30	28.25	1.31	25 to 30
Health									
Absolute									
W1	1.59	0.77	1 to 4	1.58	0.67	1 to 4	1.63	0.83	0 to 4
W2	1.80	0.81	1 to 4	1.94	0.80	1 to 5	1.94	0.72	1 to 4
Relative									
W1	1.71	0.78	1 to 4	1.84	0.72	1 to 4	1.86	0.86	0 to 5
W2	1.72	0.74	1 to 4	1.60	0.68	1 to 4	1.70	0.66	1 to 3

Note. W1 = Wave 1; W2 = Wave 2; CE = cognitively elite; CN = cognitively normal; CI = cognitively impaired; MMSE = Mini-Mental State Examination; absolute = health relative to a perfect state; relative = health relative to same-age peers.

Table 2

Means and Standard Deviations for Executive Function Tests at W1 and W2

Executive function test	CE			CN			CI		
	M	SD	Range	M	SD	Range	M	SD	Range
Hayling									
W1	5.90	1.13	3 to 8	5.80	1.26	1 to 9	5.47	1.40	1 to 8
W2	5.74	1.18	3 to 9	5.63	1.39	1 to 9	5.15	1.65	1 to 7
Stroop									
W1	1.46	0.48	-0.17 to 2.54	1.18	0.62	-1.03 to 2.52	1.15	0.72	-1.38 to 2.36
W2	1.67	0.47	0.11 to 2.43	1.29	0.89	-5.09 to 2.50	1.16	1.04	-3.15 to 2.45
Brixton									
W1	5.35	1.79	1 to 10	4.93	2.18	1 to 10	4.59	2.26	1 to 10
W2	5.73	1.75	1 to 8	5.35	2.10	1 to 10	5.29	2.21	1 to 10
Color Trails									
W1	109.67	19.48	46.70 to 140.97	94.16	27.00	-33.87 to 136.71	80.38	40.42	-78.58 to 139.37
W2	110.35	23.23	10.99 to 139.17	96.08	26.59	-21.57 to 147.09	79.55	42.14	-81.17 to 133.08
Reading Span									
W1	3.63	0.99	2 to 6	2.90	0.94	1 to 6	2.52	1.01	0 to 5
W2	3.45	0.95	1 to 5	2.71	0.93	1 to 5	2.37	0.97	0 to 4
Computational Span									
W1	4.09	1.51	1 to 7	3.12	1.13	1 to 7	2.73	1.08	0 to 6
W2	3.92	1.52	1 to 7	2.98	1.07	1 to 6	2.38	1.10	0 to 6

Note. A higher score on all six tasks indicates better executive functioning. W1 = Wave 1; W2 = Wave 2; CE = cognitively elite; CN = cognitively normal; CI = cognitively impaired.

\$watermark-text

\$watermark-text

\$watermark-text

Table 3
Zero-Order Correlations Between Executive Function Tests at Wave 1 for Each Cognitive Status Group

	Hayling	Stroop	Brixton	Color Trails	Reading Span	Computational Span
Hayling	—					
Stroop	.07, .04, .06	—				
Brixton	-.15, .18, .03	.09, .05, .03	—			
Color Trails	.19, .06, .28**	.12, .07, .29***	.07, .24***, .15	—		
Reading Span	.06, .19***, .13	.22*, .14*, .35***	.04, .18***, .19*	.19, .21***, .36***	—	
Computational Span	.06, .07, .11	.14, .12*, .28**	.11, .15***, .14	.29***, .18***, .32***	.52***, .49***, .50***	—

Note. Sample size range are cognitively elite (CE): *rs* = 77 to 80; cognitively normal (CN): *rs* = 277 to 284; cognitively impaired (CI): *rs* = 125 to 131. Correlations are listed in the following order: CE, CN, and CI.

* *p* < .05.
** *p* < .01.
*** *p* < .001.

\$watermark-text

\$watermark-text

\$watermark-text

Table 4
Descriptive Statistics and Test–Retest Correlations of Executive Function Variables

Variable	W1 <i>M</i> (<i>SD</i>)	W2 <i>M</i> (<i>SD</i>)	η^2	$r_{1,2}^b$	W1 <i>r</i> _{age}	W2 <i>r</i> _{age}	W1 <i>r</i> _{ed}	W2 <i>r</i> _{ed}
Hayling	5.81 (1.26)	5.54 (1.43)	3.41***	.43***	-.29***	-.39***	.07	.01
Stroop	1.23 (0.61)	1.34 (0.88)	-2.37*	.44***	-.21***	-.32***	.07	.02
Brixton	5.09 (2.11)	5.44 (2.05)	-2.67**	.35***	-.27***	-.33***	.20***	.18***
Color Trails	96.94 (26.87)	94.89 (31.99)	1.51	.65***	-.38***	-.49***	.13**	.10
Reading Span	3.02 (1.02)	2.77 (1.00)	4.92***	.54***	-.29***	-.41***	.22***	.15**
Computational Span	3.25 (1.25)	3.02 (1.27)	3.82***	.61***	-.19***	-.31***	.22***	.17***

Note. Values are for those persons selected into a cognitive status group at W1. Values in column 2 (W1, *M* and *SD*) refer to participants who returned at W2, as these data are used for the descriptive statistics in columns 3 and 4. W1 = Wave 1; W2 = Wave 2; ed = education.

^a *t* value of paired samples *t* test for W1 versus W2. The multivariate Group × Time interaction for the executive function variables was not significant.

^b Autocorrelations at W1 and W2. Ed = education.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 5

Standardized Factor Loadings for Executive Function Tests at Wave 1

	EF factor		
	CE	CN	CI
Hayling	.23	.09	.17
Stroop	.20	.24	.44
Brixton	.28	.11	.21
Color Trails	.28	.35	.47
Reading Span	.79	.67	.76
Computational Span	.62	.83	.67

Note. Factor loadings are derived from the one-factor models. EF = executive function; CE = cognitively elite; CN = cognitively normal; CI = cognitively impaired.

Table 6

Goodness-of-Fit Statistics for Latent Models at Wave 1 and Longitudinally

Model	χ^2	df	p	CFI	GFI	RMSEA	$\Delta\chi^2$	Δdf	p
Research Goal 1: Structure of EF (three factor)									
M1 (CE)	5.55	6	.48	.98	1.00	.00			
M2 (CN)	5.11	6	.53	1.00	.99	.00			
Structure of EF (one factor)									
M3 (CE)	6.53	9	.69	1.00	.97	.00			
M4 (CN)	17.11	9	.05	.94	.98	.06			
M5 (CI)	5.63	9	.78	1.00	.98	.00			
Research Goal 2: Group invariance (one factor)									
M6 _c (CN, CE)	23.64	18	.17	.96	.97	.04			
M7 _m (CN, CE)	31.41	23	.11	.95	.95	.05	7.77	5	>.05
M8 _c (CN, CI)	22.73	18	.20	.98	.98	.04	—	—	—
M9 _m (CN, CI)	31.81	23	.10	.96	.97	.04	9.08	5	>.05
M10 _c (CE, CI)	12.16	18	.84	1.00	.97	.00	—	—	—
M11 _m (CE, CI)	25.80	23	.31	.97	.94	.04	13.64	5	<.05
Research Goal 3: Longitudinal invariance									
M12 _c (CE, 3f)	28.43	33	.69	.98	.93	.00	—	—	—
M13 _m (CE, 3f)	37.93	39	.74	.98	.92	.00	4.62	6	>.05
M14 _c (CN, 3f)	38.28	33	.24	.97	.97	.03	—	—	—
M15 _m (CN, 3f)	48.09	39	.09	.95	.96	.04	9.81	6	>.05
M16 _c (CE, 1f)	60.71	47	.09	.92	.86	.07	—	—	—
M17 _c (CN, 1f)	67.88	47	.02	.92	.94	.05	5.14	5	<.05
M18 _m (CN, 1f)	84.23	52	.00	.95	.93	.06	16.35	5	<.05
M19 _m (CI, 1f)	55.50	47	.18	.95	.88	.05	—	—	—
M20 _m (CI, 1f)	65.11	52	.10	.94	.87	.06	9.61	5	>.05
Research Goal 4:									
M21 (stable CN, 3f)	3.64	6	.73	1.00	.99	.00			
M22 _c (stable CN, 3f)	44.80	33	.08	.98	.95	.05			

\$watermark-text

\$watermark-text

\$watermark-text

Model	χ^2	df	p	CFI	GFI	RMSEA	$\Delta\chi^2$	Δdf	p
M23 _m (stable CN, 3f)	51.70	39	.08	.97	.94	.05	6.90	6	<.05
M24 (stable CN, 1f)	8.82	9	.45	1.00	.98	.00			
M25 _c (stable CN, 1f)	61.58	47	.08	.91	.82	.08			
M26 (stable CI, 1f)	14.04	9	.12	.77	.91	.11			

Note. The three-factor model for the CI group (and stable CI group) did not converge. The three-factor model did not converge between the CE to CN group comparisons indicating a lack of group invariance. CFI = comparative fit index; GFI = goodness-of-fit index; RMSEA = root-mean-square error of approximation; EF = executive function; M = model; CE = cognitively elite; CN = cognitively normal; CI = cognitively impaired; c = configural invariance; m = metric invariance; 1f = one-factor model; 3f = three-factor model; M22, M23, M25 are longitudinal invariance models.