"Executive functions" cannot be distinguished from general intelligence: two variations on a single theme within a symphony of latent variance

Donald R. Royall[1,2,3,4]* and Raymond F. Palmer[3]

[1] Department of Psychiatry, The University of Texas Health Science Center, San Antonio, San Antonio, TX, USA
[2] Department of Medicine, The University of Texas Health Science Center, San Antonio, San Antonio, TX, USA
[3] Department of Family and Community Medicine, The University of Texas Health Science Center, San Antonio, San Antonio, TX, USA
[4] The South Texas Veterans' Health System Audie L. Murphy Division, Geriatric Research Education and Clinical Center, San Antonio, TX, USA

The empirical foundation of executive control function (ECF) remains controversial. We have employed structural equation models (SEM) to explicitly distinguish domain-specific variance in executive function (EF) performance from memory (MEM) and shared cognitive performance variance, i.e., Spearman's "g." EF does not survive adjustment for both MEM and g in a well fitting model of data obtained from non-demented older persons ($N = 193$). Instead, the variance in putative EF measures is attributable only to g, and related to functional status only through a fraction of that construct (i.e., "d"). d is a homolog of the latent variable δ, which we have previously associated specifically with the Default Mode Network (DMN). These findings undermine the validity of EF and its putative association with the frontal lobe. ECF may have no existence independent of general intelligence, and no functionally salient association with the frontal lobe outside of that structure's contribution to the DMN.

Keywords: aging, cognition, dementia, executive function, functional status, g, intelligence

INTRODUCTION

Executive Control Function (ECF) is widely thought to be vital to human autonomy, and a major determinant of problem behavior and disability in neuropsychiatric disorders (Royall et al., 2002a). Nevertheless, we lack a "gold standard" ECF measure, and the construct as a whole seems to lack a coherent empirical foundation.

"Executive functions" (EF) broadly encompass cognitive skills that are responsible for the planning, initiation, sequencing, and monitoring of complex goal-directed behavior. This may explain the relatively robust associations between EF measures and Instrumental Activities of Daily Living (IADL) (Royall et al., 2007).

However, the relationship between EF and functional status is more complex. Individual EF measures empirically load on more than one "executive" factor (Miyake et al., 2000; Royall et al., 2003; Androver-Roig et al., 2012; Testa et al., 2012). Neither the EF factor nor their indicators are necessarily associated with IADL. Executive measures are therefore commonly "validated" against structural or functional frontal lobe pathology. However, these associations are statistically weak to moderate, and qualitatively non-specific. Many executive tasks and measures can be associated with non-frontal structures and lesions (Collette and Van der Linden, 2002; Alvarez and Emory, 2006).

Recently, my colleagues and I have examined the "cognitive correlates of functional status" as a latent variable (i.e., "δ" for "dementia") in a Structural Equation Model (SEM) framework (Royall and Palmer, 2012, 2013, 2014; Royall et al., 2012a,b,

2013). δ and its homologs are strongly associated with IADL, more strongly so than are any of their indicators, including EF measures.

δ's design explicitly parses a battery's shared variance (i.e., Spearman's g) into orthogonal fractions (g' and δ) of which only δ is related to functional status (i.e., δ's "target indicator") (Royall and Palmer, 2012). δ "homologs" can be constructed from any battery that contains both cognitive measures and one or more measures of IADL.

By definition, dementia requires disabling cognitive impairment. Therefore, only δ's variance is both necessary and sufficient to dementia case finding. Thus, δ scores can be interpreted as a dementia phenotype. δ homologs have achieved Areas Under the Receiver Operating Curve (AUC /ROC) of 0.92–0.99 for the discrimination of well-characterized Alzheimer's Disease (AD) cases vs. controls in four datasets to date, although each δ homolog accounts for a minority of the variance in observed cognitive performance. The latent variable g' (δ's residual in Spearman's g) and measurement "error" (including domain specific variance) account for the majority of cognitive variance, yet g' has an AUC of only 0.52–0.66 (Royall et al., 2012a,b, 2013; Royall and Palmer, 2013, 2014). δ has been independently validated by a second group using the National Alzheimer's Coordinating Center's (NACC) Unified Dataset (UDS) (Gavett et al., 2014). In that dataset ($N \approx 26,000$), δ had an AUC of 0.96 for the discrimination between demented and non-demented participants, vs. g's 0.52. It is important to note that the NACC dataset is not limited to AD, but includes cases with a variety of dementing illnesses. This

supports δ's association with dementia in the abstract, regardless of its etiology.

δ and its homologs are derived from Spearman's general intelligence factor, "g" (Spearman, 1904), i.e., a latent variable representing the shared variance in the dominant factor extracted from any cognitive battery. The latent variable g, in turn, has been associated with frontal lobe lesions (Duncan and Owen, 2000; Duncan et al., 2000) executive measures (Duncan et al., 1997), and frontal lobe imaging (Choi et al., 2008; Gläscher et al., 2010). Since "g" can also be associated with functional outcomes (Gottfredson, 1997), we decided to explore whether an EF specific factor can be distinguished from other domain-specific variance (i.e., memory) and/or g-δ. If not, then EF may merely represent g or δ's influence on cognitive task performance, and δ may represent the emergent "ECF" responsible for uniquely human "executive" capacities.

METHODS

AIR FORCE VILLAGES' FREEDOM HOUSE STUDY

We have studied 547 well elderly retirees as part of the Air Force Villages' (AFV) Freedom House Study (FHS). The AFV is a 1500-bed Comprehensive Care Retirement Community in San Antonio, TX that is open to Air Force officers and their dependents. At baseline, the FHS subjects represented a random sample of AFV residents over the age of 70 years living at non-institutionalized levels of care. Informed consent was obtained prior to their evaluations.

A subset of FHS participants ($n = 193$) were administered a formal neuropsychological test battery that included standardized tests of memory, language, and ECF. This subgroup was slightly older at baseline than the larger FHS cohort (mean age of 79.0 years vs. 77.7 years, respectively), but did not differ significantly with regard to gender, education, baseline level of care, or Mini-Mental State Examination (MMSE) scores (Folstein et al., 1975). Select demographic and clinical features are presented in **Table 1**.

COGNITIVE BATTERY

Memory measures

The California Verbal Learning Task (CVLT) (Delis et al., 1987) assesses learning and memory processes. Patients are asked to learn and recall two 16 item shopping lists. Each list is comprised of four words from four semantic categories. Learning takes place over five trial presentations. We modeled the summed number of correct words recalled across learning trials 1–5.

The Mattis Dementia Rating Scale: memory subscale (DRS:MEM) (Mattis, 1988) provides a brief assessment of verbal and nonverbal short-term memory. The memory subtest consists of sentence (five word) recall, design and word recognition, and orientation items.

"Executive" measures

CLOX: An Executive Clock Drawing Task (Royall et al., 1998b) is a brief ECF measure based on a clock-drawing task (CDT). It is divided into two parts. CLOX1 is an unprompted task that is sensitive to executive control. CLOX2 is a copied version that is less dependent on executive skills. Each CLOX subtest is scored on a 15-point scale. Lower CLOX scores are impaired.

The Executive Interview (EXIT25) (Royall et al., 1992) provides a standardized clinical EF assessment. It contains 25 items

Table 1 | Subject characteristics.

Variable	$N = 193$
AGE (years)	77.9 (4.9)
Education (years)	15.1 (2.4)
IADL total (MAX = 14.0)	13.3 (1.7)
ADL total (MAX = 14.0)	13.5 (1.4)
MD visits previous 6 months	3.8 (4.1)
VISION (2 = "good"; 3 = "fair")	2.2 (0.8)
% Female	58.3
% Living alone	27.8
% Reporting	
Glaucoma	18.3
Arthritis	61.2
HTN	41.9
CAD	22.5
CVA	6.3
AODM	5.6
% Using prostheses	12.0

ADL's, activities of daily living; AODM, adult onset diabetes mellitus; CAD, coronary artery disease; CVA, cerebrovascular disease, IADL's, Instrumental Activities of Daily Living; HTN, hypertension; MAX, maximum; MD, physician.

designed to elicit signs of frontal system pathology (e.g., imitation, intrusions, disinhibition, environmental dependency, perseveration, and frontal release). EXIT25 scores range from 0 to 50. High scores indicate impairment.

The Controlled Oral Word Association (COWA) (Benton and Hamsher, 1989) is a test of oral word production (verbal fluency). The patient is asked to say as many words as they can, beginning with a certain letter of the alphabet.

The WAIS-R Digit Symbol Coding (DSS) (Wechsler, 1991) is a test of psychomotor speed and attentional control the subject is asked to copy as quickly as possible, nonsense symbols corresponding to specific numbers presented in a "key" at the top of the page.

The Trail Making Test, Parts A and B (Reitan, 1958) provide a measure of conceptualization, psychomotor speed, and attention. Trails B requires the subject to connect consecutively numbered and lettered circles, alternating between the two sequences.

The abbreviated Wisconsin Card Test (Haaland et al., 1987) is an adaptation of the original two deck (128 cards) Wisconsin Card Sorting Test (WCST) (Heaton et al., 1993). The Abbreviated WCST utilizes one deck of 64 cards. The number of "categories correct" (WCAT) was used as an outcome measure.

Although the above are all widely considered to be validated "executive" measures, they empirically load on at least three factors (Royall et al., 2003).

FUNCTIONAL STATUS

Disability and comorbid medical conditions were assessed using the Older Adults Resources Scale (OARS) (Fillenbaum, 1978). The OARS is a structured clinical interview that provides self-reported information on activities of daily living (ADL), IADL, physical and mental health history, healthcare utilization, and current medications.

STATISTICAL APPROACH

This analysis was performed using Analysis of Moment Structures (AMOS) software (Arbuckle, 2006). All analyses were conducted in an SEM framework.

Analysis sequence

First we examined the associations between individual cognitive performance measures and IADL in a multivariate regression model, adjusted for age, education, and gender. The covariates were entered first, and their effect on IADL established. Then the entire set of cognitive performance measures was added as predictors. IADL was used as the dependent variable. Model fit was examined.

Next, we reorganized the observed variables as a confirmatory bifactor measurement model, testing our apriori assumptions about which measures can be associated with domain specific "memory" and "executive" factors (i.e., "MEM" and "ECF," respectively). All indicators were adjusted for age, education, and gender. The relative correlations between both latent constructs and IADL were determined. Model fit was again examined.

Next, we introduced a third latent construct representing Spearman's general intelligence factor "g." The entire battery of psychometric measures was used as g's indicators. We examined the effect of g's introduction the latent domain specific factors and their indicator weights. As before, all indicators were also adjusted for age, education, and gender. The relative correlations between g, the domain specific latent constructs and IADL were determined. Model fit was again examined.

Next, we reorganized g into IADL-related and independent fractions (i.e., "d" and "g," respectively), as previously described (e.g., Royall and Palmer, 2013). By definition, g′ had no association with IADL. The relative correlations between d, the domain specific latent constructs and IADL were determined. Model fit was again examined.

Next, we searched for additional measure specific associations between individual cognitive measures and IADL, independent of the latent constructs. Finally, we systematically explored the possibility of significant intercorrelations amongst the indicator variables' residuals, which might suggest the existence of additional latent constructs other than g′, d, MEM, and EF. Only intercorrelations between two indicators' residuals that were statistically significant, improved model fit and did not result in negative variance or other model misspecifications, were retained.

Missing data

These models were all constructed in an SEM framework, using raw data. Modern Missing Data Methods were automatically applied by the AMOS software. AMOS uses Full information Maximum Likelihood (FIML) methods to address missing data. FIML uses the entire observed data matrix to estimate parameters with missing data. In contrast to list wise or pair wise deletion, FIML yields unbiased parameter estimates, preserves the overall power of the analysis, and is arguably superior to alternative methods, e.g., multiple imputation (Schafer and Graham, 2002; Graham, 2009).

Fit indices

Model fit was assessed using four common test statistics: chi-square, the ratio of the chi-square to the degrees of freedom in the model (CMIN /DF), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). Where two nested models were compared, the Browne–Cudek Criterion (BCC) was added (Browne and Cudeck, 1989).

A non-significant chi-square signifies that the data are consistent with the model (Bollen and Long, 1993). However, in large samples, this metric is limited by its tendency to achieve statistical significance when all other fit indices (which are not sensitive to sample size) show that the model fits the data very well. A CMIN/DF ratio <5.0 suggests an adequate fit to the data (Wheaton et al., 1977). The CFI statistic compares the specified model with a null model (Bentler, 1990). CFI values range from 0 to 1.0. Values below 0.95 suggest model misspecification. Values approaching 1.0 indicate adequate to excellent fit. An RMSEA of 0.05 or less indicates a close fit to the data, with models below 0.05 considered "good" fit, and up to 0.08 as "acceptable" (Browne and Cudeck, 1993). A lower BCC statistic indicates better fit (Browne and Cudeck, 1989). All fit statistics should be simultaneously considered when assessing the adequacy of the models to the data.

RESULTS

Sample demographics are presented in **Table 1**. Clinical assessment means are presented in **Table 2**. Model 1's fit was poor (**Table 3**). Together, the cognitive performance measures and covariates explained 24.1% of variance in IADL. Age, gender, DSS ($r = 0.224$, $p = 0.001$) DRS:MEM ($r = 0.158$, $p = 0.02$) and EXIT25 (partial $r = -0.145$, $p < 0.001$), contributed significantly to IADL, similar to previous analyses in this cohort (Royall et al., 2000, 2004, 2005a,b) (**Figure 1**).

Model 2 posits two domain specific factors, MEM and EF (**Figure 2**). The fit of this model is significantly improved relative

Table 2 | Raw cognitive performance means.

Variable ($N = 193$)	Mean (SD)
MEMORY TESTS	
CVLT: 1–5	32.8 (14.1)
CVLT: Long	16.0 (3.7)
CVLT: Short	15.0(3.8)
DRS: MEM	21.5 (4.3)
ECF TESTS	
CLOX1	10.1 (3.3)
COWA	32.0 (12.6)
DSS	33.2 (11.6)
EXIT25	14.6 (5.5)
Trails B (s)	132.8 (80.8)
WCAT	2.0 (2.0)

COWA, Controlled Oral Word Association Test; CLOX1, clock drawing to command; CVLT, California Verbal Learning Test, List A delayed recall; DRS: MEM, Mattis Dementia Rating Scale Memory subscale; DSS, WAIS-R Digit Symbol Substitution; EXIT25, Executive Interview; Trails B, Trail-making Test Part B; WCAT, Wisconsin Card Sorting Test categories achieved.

Table 3 | Model fit.

Model	χ^2 (*df*), *p*	CFI	RMSEA	BCC
1	881.48 (47), $p < 0.001$	0.323	0.169	1029.03
2	142.15 (45), $p < 0.001$	0.921	0.059	293.80
3a	73.06 (39), $p = 0.001$	0.972	0.037	237.00
3b	53.99 (41), $p = 0.08$	0.989	0.023	213.83
4	36.12 (31), $p = 0.24$	0.996	0.016	216.46

*BCC, Browne–Cudek Criterion; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation.*

to Model 1 (**Table 3**). CVLT:Short, CVLT:Long, CVLT 1–5, and DRS MEM all load significantly on MEM (all $p < 0.001$). The strengths of their loadings ranged from $r = 0.52$ (DRS:MEM) to $r = 0.90$ (CVLT:Long). CLOX1, DSS, EXIT25, Trails B, and WCAT all load significantly on EF (all $p = 0.002$). The strengths of their loadings ranged from $r = -0.25$ (Trails B) to 0.66 (DSS). MEM and EF were uncorrelated. As expected, EF was significantly associated with IADL ($r = 0.34$, $p < 0.001$). MEM was weakly but significantly correlated with IADL independent of EF ($r = 0.17$, $p = 0.02$).

Model 3 posited the addition of a third factor, Spearman's g. Our first attempt at a three factor model failed (due to unsuccessful minimization and negative variance). Minimization could be achieved by correlating EF and MEM but (1) the correlation between MEM and EF was not significant ($r = -0.04$, $p = 0.923$), (2) negative variance persisted on COWA's residual, (3) EF lost its association with IADL ($r = -0.05$, $p = 0.924$), (4) EF had no significant indicators (all $p > 0.92$).

Two alternative two factor models were then tested. Model 3a omitted the factor MEM (**Table 3**). Model 3b omitted the factor EF. In each case, these models containing g fit the data better than Models 1 or 2. In each case, the latent variable g had a stronger correlation with IADL than did the second domain specific factor. In Model 3a, g fully mediated EF's previously significant association with IADL in Model 2. However, Model 3b fit the data significantly better than did Model 3a. On the basis of these findings, the latent factor EF was deleted from subsequent models.

In the adopted Model 3b (**Figure 3**), g was indicated significantly by all the cognitive measures (all $p \leq 0.002$) ranging from Trails B ($r = -0.23$, $p = 0.002$) to DSS ($r = 0.66$, $p < 0.001$). MEM's factor loadings were slightly attenuated by g's creation, ranging from $r = 0.23$ (DRS:MEM) to $r = 0.70$ (CVLT:Long). g was significantly correlated with IADL ($r = 0.40$, $p < 0.001$). MEM had no significant association with that variable ($r = 0.09$, $p = 0.261$). Thus, g both mediates MEM's unadjusted association with IADL and better fits the variance in our putative ECF measures than would an EF domain-specific factor, whether adjusted for g or not.

Model 4 parses Spearman's g into two fractions (**Figure 3**). d is indicated by IADL and the cognitive performance measures. g′ (i.e., d's residual in Spearman's g) and MEM are indicated only by cognitive performance measures. This arrangement had excellent fit, and fit the data significantly better than any previous model (**Table 3**). d was significantly indicated by all the
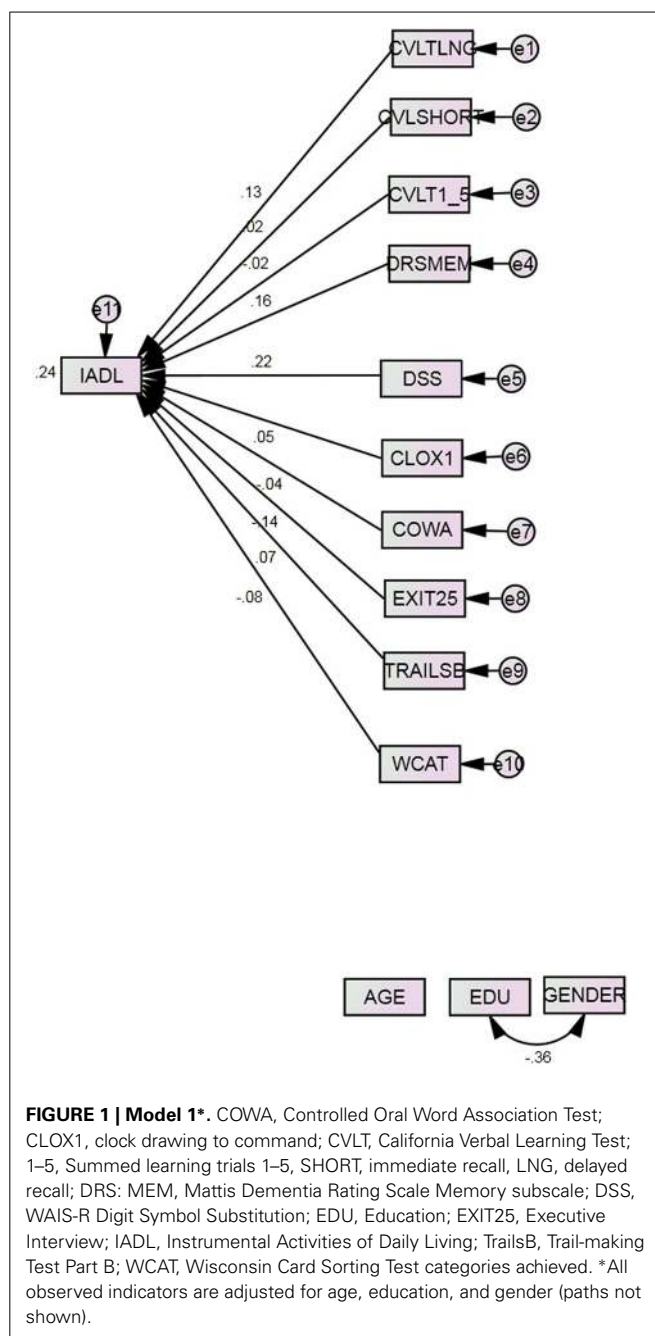
FIGURE 1 | Model 1*. COWA, Controlled Oral Word Association Test; CLOX1, clock drawing to command; CVLT, California Verbal Learning Test; 1–5, Summed learning trials 1–5, SHORT, immediate recall, LNG, delayed recall; DRS: MEM, Mattis Dementia Rating Scale Memory subscale; DSS, WAIS-R Digit Symbol Substitution; EDU, Education; EXIT25, Executive Interview; IADL, Instrumental Activities of Daily Living; TrailsB, Trail-making Test Part B; WCAT, Wisconsin Card Sorting Test categories achieved. *All observed indicators are adjusted for age, education, and gender (paths not shown).

cognitive measures except WCAT ($r = 0.10$, $p = 0.30$) and Trails B ($r = 0.05$, $p = 0.63$). WCAT and Trails B loaded significantly on g′ (both $p \leq 0.002$) as did all the other cognitive measures, ranging from CLOX1 ($r = -0.27$) to COWA ($r = -0.62$, both $p < 0.001$).

However, by definition, g′ had no association with IADL. In contrast, d was associated strongly with IADL ($r = 0.65$, $p < 0.001$). Independently of their associations with d, no cognitive performance measure was significantly associated with IADL, i.e., through their residuals. Thus, WCAT and Trails B had no significant associations with IADL at all.
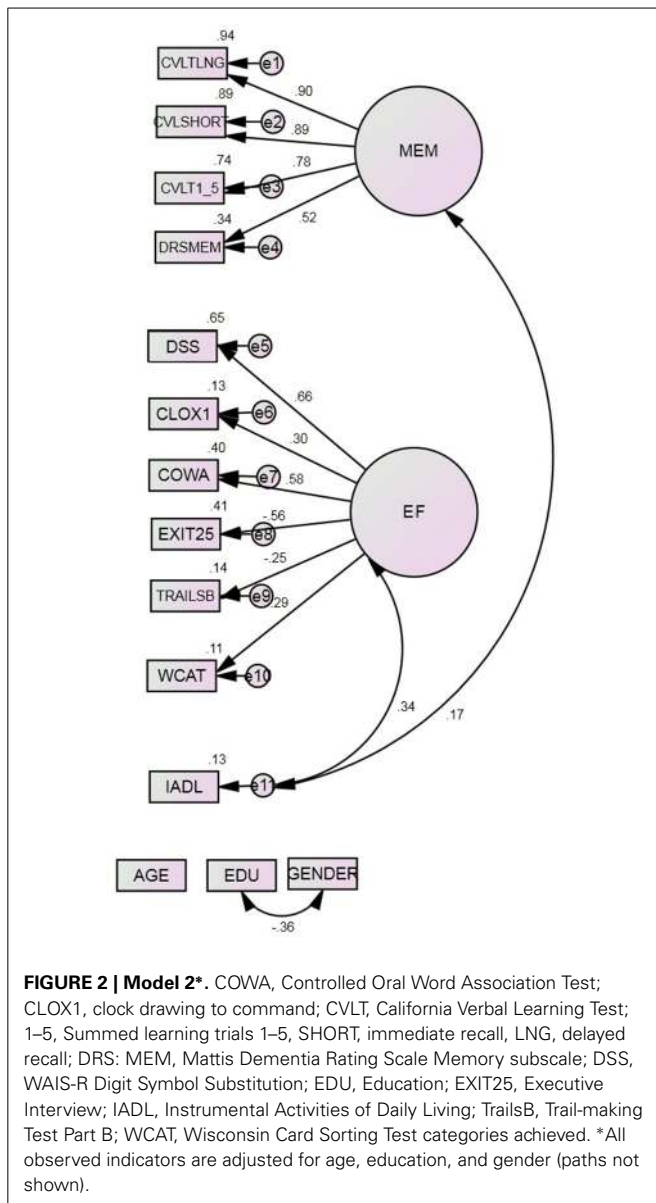
FIGURE 2 | Model 2*. COWA, Controlled Oral Word Association Test; CLOX1, clock drawing to command; CVLT, California Verbal Learning Test; 1–5, Summed learning trials 1–5, SHORT, immediate recall, LNG, delayed recall; DRS: MEM, Mattis Dementia Rating Scale Memory subscale; DSS, WAIS-R Digit Symbol Substitution; EDU, Education; EXIT25, Executive Interview; IADL, Instrumental Activities of Daily Living; TrailsB, Trail-making Test Part B; WCAT, Wisconsin Card Sorting Test categories achieved. *All observed indicators are adjusted for age, education, and gender (paths not shown).

There were no significant intercorrelations amongst the residuals of the final three latent constructs' indicators, in Model 4. Specifically, none of the ECF measures' residuals were significantly correlated. This finding closes the door to the possibility of one or more unmodeled factors, including EF or processing speed. Since the modeled factors explain a minority of the variance in most ECF measures (**Figure 4**), their uncorrelated residuals may reflect measure specific "measurement error." By definition, the three latent variables d, g′, and MEM were orthogonal to each other and could not be intercorrelated.

DISCUSSION

In this analysis, we have confirmed the relatively strong association between the EXIT25, and IADL, in a multivariate regression model. The EXIT25 contributed significantly to IADL

independent of memory measures and a battery of other EF measures. This is consistent with several previous studies in a wide range of samples (Chan et al., 2006; Lewis and Miller, 2007; Pereira et al., 2008), including this one (Royall et al., 1998a, 2000, 2004, 2005a,b).

Together, cognitive measures and covariates explained a respectable fraction of IADL variance. However, the model did not fit the data well. The SEM approach forces our attention to the quality of a model's fit, not merely the significance of its parameters and the total variance explained in its dependent variable. In every case, the introduction of latent variables fit the data better than did our initial multivariate regression approach.

Model 2 has confirmed our apriori assumptions about the domain specific face validity of our cognitive battery. All the memory measures loaded significantly on the latent construct "MEM." All the executive measures loaded significantly on the latent construct "EF." These factors were not significantly associated with each other. As we expected, EF was more strongly associated with IADL than MEM, which was weakly associated with that construct.
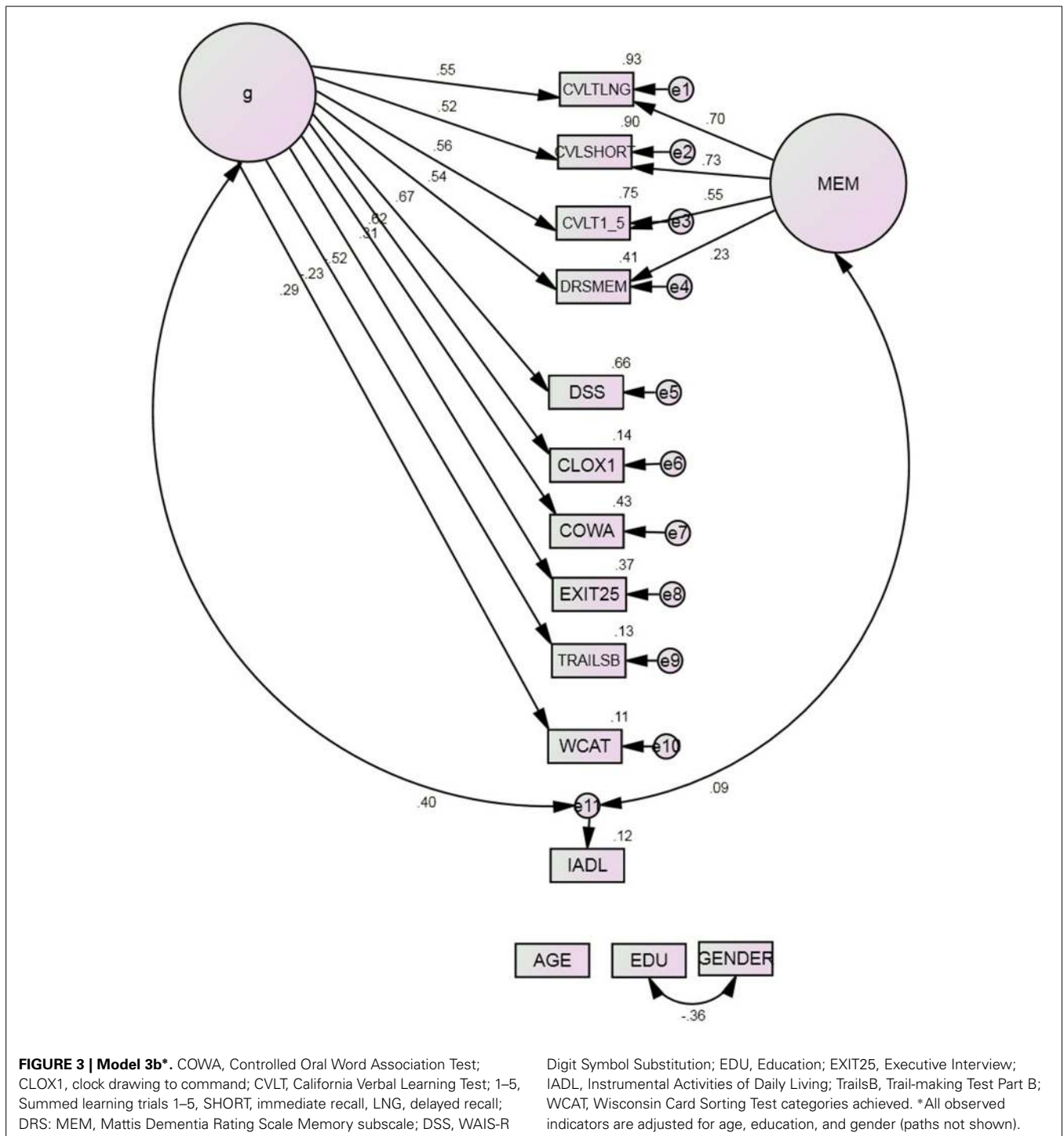
However, subsequent models with better fit have forced us to abandon the EF construct. The introduction of g′ and d provide a much better fit, and the absence of significant intercorrelations among their indicators' residuals closes the door to the possibility of unmodeled alternative factors (e.g., processing speed, etc.).

Models 3b and 4 suggest that EF measures have no association with IADL independent of general intelligence and specifically its subfraction δ. Model 4 demonstrates that EF measures have no special or unique association with IADL, even through d. d is also indicated by memory tasks, and they load more strongly on d than any executive measure.

Independent of d, EF measures cannot be associated with IADL, either individually (through their residuals), or via g′ (by definition). WCAT and Trails B load only on g′ and thus have no association with IADL at all. This is consistent with their failure to contribute significantly to IADL independently of the EXIT25 and other EF measures in Model 1.

Both findings also replicate our earlier factor analysis in this dataset (Royall et al., 2003). In that analysis, the variance in a battery of EF measures was empirically distributed across three factors. The first (28% of variance) was indicated by CLOX, COWA, DSS, and the EXIT25. The second (24.2% of variance) was uniquely indicated by the WCST and its subtasks. The third (12.4% of variance) was indicated uniquely by Trails B. Only the first factor was associated with IADL. The fact that d and g explain so little of the variance in our battery of otherwise non-correlated measures suggests that each EF measure may have considerable "measurement error" associated with it.

Duncan and others have previously associated g with frontal structure and function (Duncan and Owen, 2000; Duncan et al., 2000; Choi et al., 2008; Gläscher et al., 2010). Similarly, several of our EF measures have been associated with frontal structure and /or function (Royall et al., 2007; Royall, 2011). However, Model 4 demonstrates that the variance in our EF indicators is distributed across two orthogonal latent factors, d and g′. Neither is specifically associated with EF, as both are significantly indicated by also by memory tests. It is an empirical question which,
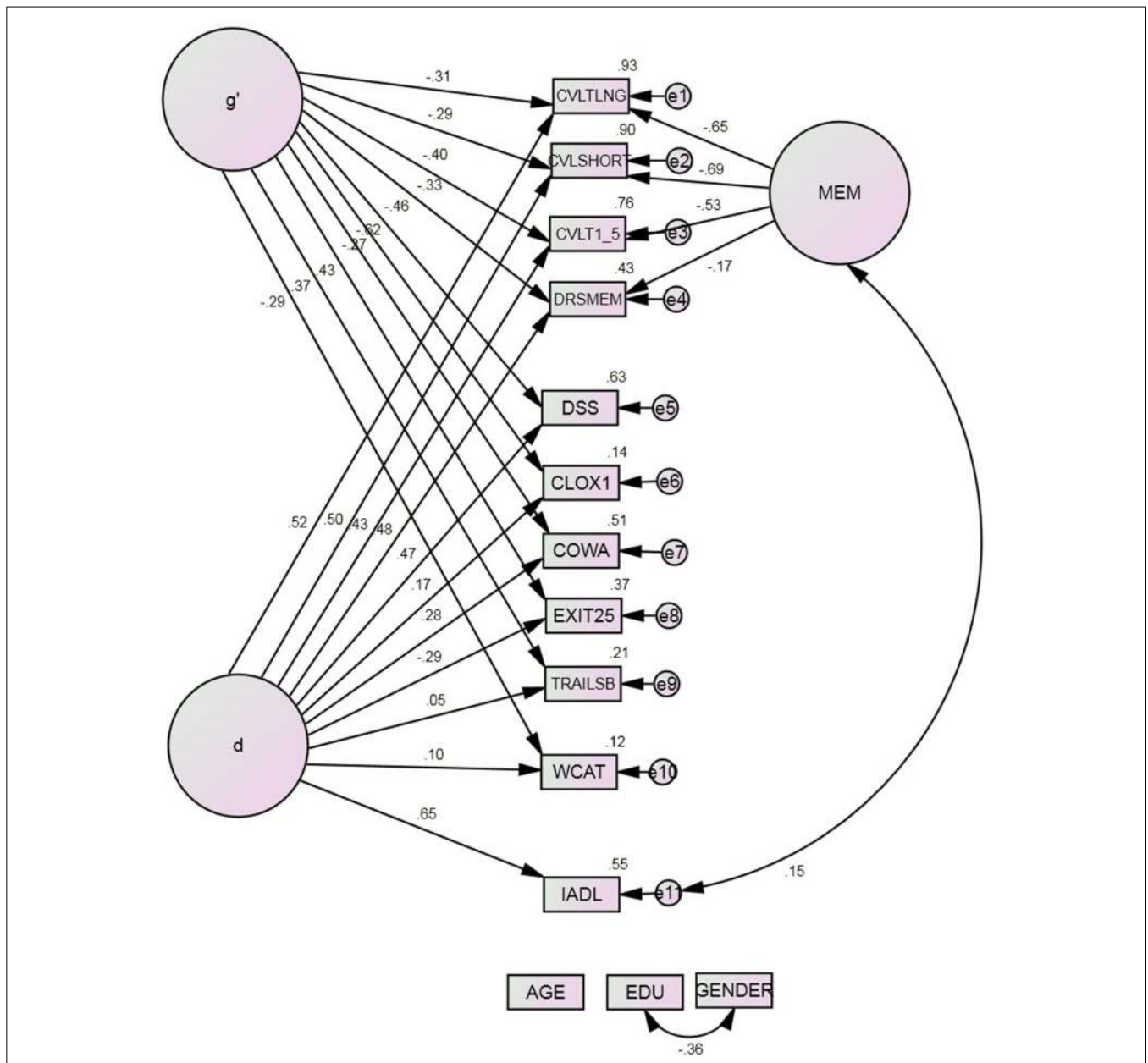
FIGURE 3 | Model 3b*. COWA, Controlled Oral Word Association Test; CLOX1, clock drawing to command; CVLT, California Verbal Learning Test; 1–5, Summed learning trials 1–5, SHORT, immediate recall, LNG, delayed recall; DRS: MEM, Mattis Dementia Rating Scale Memory subscale; DSS, WAIS-R Digit Symbol Substitution; EDU, Education; EXIT25, Executive Interview; IADL, Instrumental Activities of Daily Living; TrailsB, Trail-making Test Part B; WCAT, Wisconsin Card Sorting Test categories achieved. *All observed indicators are adjusted for age, education, and gender (paths not shown).

if either latent construct can mediate g's observed association with frontal structure and /or function. Our dataset cannot address that question.

Because g (Model 3), g' and d (Model 4) have been adjusted for memory-specific task performance (i.e., MEM), it could be argued that the loadings of memory tasks on the first three latent constructs reflects the "executive" fraction of those measures' variance (e.g., "Working Memory"). Working Memory has been

related to "updating" and can be associated with measures of intelligence (Friedman et al., 2008).

However, only d is associated with IADL. Working Memory has previously been associated with IADL (Lewis and Miller, 2007) and d is more strongly indicated by memory tasks than by executive ones. Moreover, d and g' are orthogonal to each other. Thus, they cannot both be "executive," and if g' were to be identified as the true executive factor (after all, it is most strongly loaded

FIGURE 4 | Model 4*. COWA, Controlled Oral Word Association Test; CLOX1, clock drawing to command; CVLT, California Verbal Learning Test; 1–5, Summed learning trials 1–5, SHORT, immediate recall, LNG, delayed recall; DRS: MEM, Mattis Dementia Rating Scale Memory subscale; DSS, WAIS-R Digit Symbol Substitution; EDU, Education; EXIT25, Executive Interview; IADL, Instrumental Activities of Daily Living; TrailsB, Trail-making Test Part B; WCAT, Wisconsin Card Sorting Test categories achieved. *All observed indicators are adjusted for age, education, and gender (paths not shown).

by COWA and the only factor associated with Trails B and WCAT) then EF can again have no impact on IADL.

d uniquely accounts for a sizable fraction of IADL's variance, and explains more variance in IADL than did the ECF factor in Model 2, or indeed the entire battery in Model 1. d is a homolog of δ, our latent dementia proxy. δ and its homologs are strongly and specifically associated with clinical dementia status, as measured by the Clinical Dementia Rating Scale (Hughes et al., 1982; Royall et al., 2012a,b; Royall and Palmer, 2013, 2014). Even in this non-demented cohort, the interindividual variance in δ scores predicts

longitudinal change in ECF measures, specifically the EXIT25 and Trails B (but neither WCAT nor the CVLT) (Royall and Palmer, 2012). The fact that δ predicts longitudinal change in Trails B in this very cohort suggests that Trails B's failure to load on d in this analysis may be an artifact of its baseline distribution, which is skewed and subject to floor effects. In longitudinal analyses, each subject is its own control.

In contrast to Spearman's g, δ has been associated with atrophy in the Default Mode Network (DMN) (Royall et al., 2012b, 2013). The DMN is associated with a subregion of the frontal lobe (i.e., a

small portion of the dorsolateral prefrontal cortex), but also with subregions of the temporal lobe, the parietal lobe, the cingulate gyrus and the hippocampus (Buckner et al., 2008). The latter may explain the relatively strong loadings of memory measures on d. Thus, it seems unlikely that d would localize to the frontal cortex, as might be expected of an "executive" construct (although specific frontal localizations have in fact not been shown for many executive measures).

In short, the associations between the EF factor, or its indicators and IADL are mediated uniquely through d, i.e., a fraction of Spearman's g. This result is similar to an analysis by Salthouse et al. (1996) of age's influence on cognitive task performance. They found moderately strong age-related declines on a battery of tests that included the WCST, Trails-B, and DSS, among others. However, correlation-based analyses revealed that the age-related effects on different measures were not independent. Instead, the effect of age was observed specifically in the fraction of variance (averaging 58%) shared across all the observed measures (i.e., "g"). Thus, g′ and δ may also mediate age-specific effects on ECF measures. This would explain age's broad effects on cognitive performance, relatively strong effects on "ECF" measures, and the disabling character of those effects (if mediated through δ).

On the other hand, aging is also characterized by a "de-differentiation" of cognitive test performance (McArdle et al., 2002). This may favor the demonstration of global constructs such as g, g′, and δ. It remains to be seen if a δ homolog would mediate the association(s) between one or more EF factors and IADL in healthy younger adults. One potential obstacle to such a study would be selection of a valid IADL measure. The informant-rated IADL measure we used here may have floor effects in highly functioning populations. Nevertheless, δ is not very sensitive to its IADL target, and has similar psychometric properties regardless of the target IADL indicators used to date (Royall and Palmer, 2012, 2013, 2014; Royall et al., 2012a,b; Gavett et al., 2014).

Our dataset is further limited by other issues. It does not contain measures of supposedly fundamental executive tasks (i.e., inhibition, categorization, and updating) (Miyake et al., 2000). Such measures are arguably less prone to measurement error than the "complex" ECF measures we have employed. They, and other executive tasks (e.g., set-shifting and delayed matching to sample) have been associated with frontal lobe lesions and structures. However, such low level cognitive abilities (which can be demonstrated in chimpanzees for example, at an estimated three year old human intelligence equivalent) (Moriguchi et al., 2011) may not be representative of the emergent ECF that characterizes adult human action.

It is arguable that δ cannot be demonstrated in any animal that is incapable of IADL (by definition). This may have a biological explanation. The human brain, uniquely among primates, exhibits frontal networks that extend beyond the frontal lobe (including the DMN). The frontal networks of other primates are localized to that structure (Wey et al., 2013). Frontal tasks not-related to IADL, and /or demonstrable in animals incapable of IADL, are arguably not "executive" but merely "frontal." They may be associated with δ in humans, but then so might any cognitive performance measure, whether executive or non-executive, and whether localizable to the frontal lobe or not. Regardless,

their demonstration and functional localization to frontal structures in animals incapable of IADL will not be associated with δ, by definition.

Friedman et al. (2008) have demonstrated the existence of a latent "Common" EF factor, that is indicated by all basic EF measures, as are g, g′, and δ/d. Friedman et al. distinguished their Common factor from both intelligence and processing speed. However, they did not try to associate their Common factor with non-executive indicators, and so its specificity to EF is undemonstrated, as is its association with IADL, and therefore δ.

Ironically, the Common factor's independence from intelligence suggests that it may indeed be more likely to correspond to d in this analysis than to g′, as g′ would be expected to correlate more strongly with observed performance on intelligence measures. Friedman et al. also observed that a theorized "Inhibition" factor collapsed after the Common factor's introduction. That is consistent with EF's collapse in our analysis after the introduction of g.

Second, our battery is limited in its ability to assess "processing speed." Trails B is our only timed test, although some authors associate performance on the DSS with this construct. This limits our ability to speak to processing speed as a determinant of IADL. However, such a factor is unlikely to attenuate δ's association with IADL because processing speed is an intermediate "domain-specific" factor (like MEM and EF in this analysis) and thus taps a compartment of variance in cognitive performance that is orthogonal to g (and therefore both g′ and δ). Had our battery been better designed to assess processing speed, we expect it would have robbed MEM of its relatively weak association with IADL rather than d.

Finally, this analysis is limited to cross-sectional data. At baseline, the FHS cohort was cognitively normal for its age, relatively highly functioning and non-institutionalized. Few subjects can be expected to have been clinically demented, although a sizeable fraction might have had "mild" neurocognitive disorders. Thus, restricted range and floor effects on some measures may have affected our analysis.

Clinical dementia status was never formally adjudicated in this cohort. Never the less, we have demonstrated that there is significant variability with regard to the cohort's longitudinal rates of change in cognitive performance over time (Royall et al., 2005a). These changes are clearly related to concurrent declines in functional status (Royall et al., 2004) suggesting aging-related declines in δ-specific variance. In fact, we have shown those associations to be mediated through δ (Royall and Palmer, 2012). Gavett et al. (2014) report that the six-year prospective longitudinal change in δ scores (Δδ) correlates strongly ($r = -0.94$, $p < 0.001$) with change in dementia severity, as rated by the Clinical Dementia Rating scale (CDR) (Hughes et al., 1982). Similarly, in the Texas Alzheimer's Research and Care Consortium (TARCC), δ's intercept and slope explain 79% of the variance in four year prospective dementia severity, independently of baseline dementia severity, g′ and Δg′ [Palmer and Royall (ICAAD abstract), 2013]. If ECF (as distinct from EF) is synonymous with δ then it likely is the major cognitive determinant of dementia status in humans and dementia, in turn, may be limited to structural

and functional pathologies of the DMN (Royall et al., 2002b, 2012b).

In summary, we have used a latent variable approach in an SEM framework to construct a well fitting model that suggests that the variance in a battery of well validated "executive" measures cannot be related to a domain specific "executive" factor independent of Spearman's general intelligence factor, g. Moreover, no cognitive performance measure in our battery can be associated with IADL independently of a certain fraction of that latent construct, i.e., d. d, as a δ homolog, is likely to be associated specifically with the structure and function of the DMN. That network extends well beyond the frontal lobe, and can be related only to certain subregions in that structure. This underscores the importance of disentangling "EF" from "frontal function" (Royall et al., 2002a).

Although we again confirm that memory specific variance has no association with IADL (and by extension with dementia), memory performance measures do contribute significantly to g (as should all cognitive performance measures) and its subparts: g′ and d. Only their contributions to d would be salient to functional outcomes and dementia. However, memory tasks are more strongly associated with that construct than were most "ECF" measures. It is the distribution of memory task performance across three latent constructs, two of which are irrelevant to IADL and dementia case finding that weakens their performance as predictors of IADL. In contrast, a larger share, if not the majority of variance in most putative "ECF" measures (but neither Trails B nor WCAT), is invested in δ. This explains the relatively strong associations between putative "ECF" measures, IADL and dementia status in past studies. Regardless, δ homologs should have even greater potential for dementia case-finding, although they are neither indicated solely by ECF measures, nor likely to localize specifically to the frontal lobe.

REFERENCES

Alvarez, J. A., and Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol. Rev.* 16, 17–42. doi: 10.1007/s11065-006-9002-x

Androver-Roig, D., Sesé, A., Barceló, F., and Palmer, A. (2012). A latent variable approach to executive control in healthy aging. *Brain Cogn.* 78, 284–299. doi: 10.1016/j.bandc.2012.01.005

Arbuckle, J. L. (2006). *Analysis of Moment Structures-AMOS (Version 7.0) [Computer Program]*. Chicago: SPSS.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

Benton, A., and Hamsher, K. (1989). *Multilingual Aphasia Examination*. Iowa City, IA: AJA Associates.

Bollen, K. A., and Long, J. S. (1993). *Testing Structural Equation Models*. Thousand Oaks, CA: Sage Publications.

Browne, M., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models Sage Publications*, eds K. A. Bollen and J. S. Long (Thousand Oaks, CA: Sage Publications), 136–162.

Browne, M. W., and Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivar. Behav. Res.* 24, 445–455.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The Brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Chan, S. M., Chiu, F. K., and Lam, C. W. (2006). Correlational study of the Chinese version of the executive interview (C-EXIT25) to other cognitive measures in a psychogeriatric population in Hong Kong Chinese. *Int. J. Geriatr. Psychiatry* 21, 535–541. doi: 10.1002/gps.1521

Choi, Y. Y., Shamosh, N. A., Cho, S. H., DeYoung, C. G., Lee, M. J., Lee, J.-M., et al. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. *J. Neurosci.* 28, 10323–10329. doi: 10.1523/JNEUROSCI.3259-08.2008

Collette, F., and Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neurosci. Neurobehav. Rev.* 26, 105–125. doi: 10.1016/S0149-7634(01)00063-X

Delis, D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. (1987). *California Verbal Learning Test: Adult Version Manual*. San Antonio, TX: The Psychological Corporation.

Duncan, J., Johnson, R., Swales, M., and Freer, C. (1997). Frontal lobe deficits after head injury, unity and diversity of function. *Cogn. Neuropsychol.* 14, 713–741.

Duncan, J., and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483. doi: 10.1016/S0166-2236(00)01633-7

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000). A neural basis for general intelligence. *Science* 289, 457–460. doi: 10.1126/science.289.5478.457

Fillenbaum, G. G. (1978). *Validity and Reliability of the Multidimensional Functional Assessment Questionnaire. In The OARS Methodology. Duke University Center for the Study of Aging and Human Development*. Durham, NC: Duke University.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state, a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 198–198.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., and Hewitt, J. K. (2008). *J. Exp. Psychol. Gen.* 137, 201–225. doi: 10.1037/0096-3445.137.2.201

Gavett, B. E., Vudy, V., Jeffrey, M., John, S. E., Gurnani, A., and Adams, J. (2014). *The δ latent dementia phenotype in the NACC UDS: cross-validation and extension. Neuropsychology*. doi: 10.1037/neu0000128. [Epub ahead of print].

Gläscher, J., Rudrauf, D., Colome, R., Paula, L. K., Tranelc, D., Damasio, H., et al. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4705–4709. doi: 10.1073/pnas.0910397107

Gottfredson, L. S. (1997). Why g matters, The complexity of everyday life. *Intelligence* 24, 79–132.

Graham, J. W. (2009). Missing data analysis, making it work in the real world. *Ann. Rev. Psychol.* 6, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

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

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., and Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual-Revised and Expanded*. Odessa, IL: Psychological Assessment Resources.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., and Martin, R. L. (1982). A new clinical scale for the staging of dementia. *Br. J. Psychiatry* 140, 566–572.

Lewis, M. S., and Miller, L. S. (2007). Executive control functioning and functional ability in older adults. *Clin. Neuropsychol.* 21, 274–285. doi: 10.1080/13854040500519752

Mattis, S. (1988). *Dementia Rating Scale, Professional Manual*. Odessa, FL: Psychological Assessment Resources.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., and Woodcock, R. W. (2002). Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities across the life span. *Dev. Psychol.* 38, 115–142. doi: 10.1037/0012-1649.38.1.115

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wagner, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Moriguchi, Y., Tanaka, M., and Itakura, S. (2011). Executive function in young children and chimpanzees (Pan troglodytes): evidence from a nonverbal dimensional change card sort task. *J. Genet. Psychol.* 172, 252–265. doi: 10.1080/00221325.2010.534828

Palmer, R. F., and Royall, D. R. (2013). "Future Dementia status is almost entirely explained by the latent variable "d"s baseline and change," in *Alzheimer's Association International Conference (AAIC)* (Boston, MA).

Pereira, F. S., Yassuda, M. S., Oliveira, A. M., and Forlenza, O. V. (2008). Executive dysfunction correlates with impaired functional status in older adults with varying degrees of cognitive impairment. *Int. Psychogeriatr.* 20, 1104–1115. doi: 10.1017/S1041610208007631

Reitan, R. M. (1958). Validity of the Trail Making test as an indicator of organic brain damage. *Percept. Mot. Skills* 8, 271–276.

Royall, D. R. (2011). "The executive interview," in *The Encyclopedia of Clinical Neuropsychology, 1st Edn.,* Part V, eds J. Kreutzer, J. DeLuca, and B. Caplan (New York, NY: Springer), 992–997.

Royall, D. R., Cabello, M., and Polk, M. J. (1998a). Executive dyscontrol, An important factor affecting the level of care received by elderly retirees. *J. Am. Geriatr. Soc.* 46, 1519–1524.

Royall, D. R., Chiodo, L. K., and Polk, M. J. (2000). Correlates of disability among elderly retirees with "sub-clinical" cognitive impairment. *J. Gerontol. Med. Sci.* 55A, M541–M546.

Royall, D. R., Chiodo, L. K., and Polk, M. (2003). "Executive dyscontrol in normal aging, normative data, factor structure, and clinical correlates," in *Current Neurology and Neuroscience Reports*, Vol. 3, eds J. C. M. Brust and S. Fahn (Philadelphia, PA: Current Science Inc.), 87–493.

Royall, D. R., Cordes, J. A., and Polk, M. (1998b). CLOX, An executive clock drawing task. *J. Neurol. Neurosurg. Psychiatr.* 64, 588–594.

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., et al. (2002a). Executive Control Function, a review of its promise and challenges to clinical research. *J. Neuropsychiatr. Clin. Neurosci.* 14, 377–405. doi: 10.1176/appi.neuropsych.14.4.377

Royall, D. R., Lauterbach, E. C., Kaufer, D. I., Malloy, P., Coburn, K. L., and Black, K. J. (2007). The cognitive correlates of functional status, A review from the Committee on Research of the American Neuropsychiatric Association. *J. Neuropsychiatr. Clin. Neurosci.* 19, 249–265. doi: 10.1176/appi.neuropsych.19.3.249

Royall, D. R., Mahurin, R. K., and Gray, K. F. (1992). Bedside assessment of executive cognitive impairment, The Executive Interview (EXIT). *J. Am. Geriatr. Soc.* 40, 1221–1226.

Royall, D. R., Palmer, R., Chiodo, L. K., and Polk, M. (2005a). Normal rates of cognitive change in successful aging, The Freedom House Study. *J. Int. Neuropsychol. Soc.* 11, 899–909. doi: 10.1017/S135561770505109X

Royall, D. R., Palmer, R., Chiodo, L. K., and Polk, M. J. (2004). Declining executive control in normal aging predicts change in functional status, The Freedom House Study. *J. Am. Geriatr. Soc.* 52, 346–352. doi: 10.1111/j.1532-5415.2004.52104.x

Royall, D. R., Palmer, R., Chiodo, L. K., and Polk, M. J. (2005b). Executive control mediates memory's association with change in functional status, The Freedom House Study. *J. Am. Geriatr. Soc.* 53, 11–17. doi: 10.1111/j.1532-5415.2005.53004.x

Royall, D. R., Palmer, R., Mulroy, A. R., Polk, M. J., Román, G. C., David, J. P., et al. (2002b). Pathological determinants of clinical dementia in Alzheimer's disease. *Exp. Aging Res.* 28, 143–162. doi: 10.1080/03610730252800166

Royall, D. R., and Palmer, R. F. (2012). Getting Past "g": testing a new model of dementing processes in non-demented persons. *J. Neuropsychiatr. Clin. Neurosci.* 24, 37–46. doi: 10.1176/appi.neuropsych.11040078

Royall, D. R., and Palmer, R. F. (2013). Validation of a latent construct for dementia case-finding in Mexican-Americans. *J. Alzheimers Dis.* 37, 89–97. doi: 10.3233/JAD-130353

Royall, D. R., and Palmer, R. F. (2014). Does ethnicity moderate dementia's biomarkers? *Neurobiol. Aging* 35, 336–344. doi: 10.1016/j.neurobiolaging.2013.08.006

Royall, D. R., Palmer, R. F., and O'Bryant, S. E. (2012a). Validation of a latent variable representing the dementing process. *J. Alzheimers Dis.* 30, 639–649. doi: 10.3233/JAD-2012-120055

Royall, D. R., Palmer, R. F., Vidoni, E. D., and Honea, R. A. (2013). The default mode network may be the key substrate of depression-related cognitive changes. *J. Alzheimers Dis.* 34, 547–560. doi: 10.3233/JAD-121639

Royall, D. R., Palmer, R. F., Vidoni, E. D., Honea, R. A., and Burns, J. M. (2012b). The Default Mode Network and related right hemisphere structures may be the key substrates of dementia. *J. Alzheimers Dis.* 32, 467–478. doi: 10.3233/JAD-2012-120424

Salthouse, T. A., Hancock, H. E., and Meinz, E. J. (1996). Interrelations of age, visual acuity, and cognitive functioning, *J. Gerontol. B Psychol. Sci.* 51, P317–P330.

Schafer, J. L., and Graham, J. W. (2002). Missing data, our view of the state of the art. *Psychol. Methods* 7, 147–177. doi: 10.1037/1082-989X.7.2.147

Spearman, C. (1904). General intelligence, objectively determined and measured. *Am. J. Psychol.* 15, 201–293.

Testa, R., Bennett, P., and Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Arch. Clin. Neuropsychol.* 27, 213–224. doi: 10.1093/arclin/acr112

Wechsler, D. (1991). *Wechsler Adult Intelligence Scale-Revised*. New York, NY: Psychological Corporation.

Wey, H. Y., Phillips, K. A., McKay, D. R., Laird, A. R., Kochunov, P., Davis, M. D., et al. (2013). Multi-region hemispheric specialization differentiates human from nonhuman primate brain function. *Brain Struct. Funct.* doi: 10.1007/s00429-013-0620-9. [Epub ahead of print].

Wheaton, B., Muthen, B., Alwin, D. F., and Summers, G. F. (1977). "Assessing reliability and stability in panel models with multiple indicators," in *Sociological Methodology*, ed D. R. Heise (San Francisco, CA: Jossey-Bass), 84–136.