Supplementary Materials

1. Participant Exclusion. The NKI-RS excluded participants if they were diagnosed with or reported severe psychiatric disorders (bipolar disorder, schizophrenia disorder, schizoaffective disorder), severe developmental disorders (autism spectrum disorders, intellectual disabilities), current suicidal or homicidal ideation, severe cerebral trauma (stroke, moderate to severe traumatic brain injury, transient ischemic attack in the past two years), severe neurodegenerative disorders (Parkinson's disease, Huntington's disease, dementia), a history of substance dependence in the past two years (except cannabis), a lifetime history of psychiatric hospitalization, current pregnancy, or MRI contraindications. Participants were excluded from these analyses if T1-weighted structural images were incomplete or were unable to be reconstructed due to significant image artifacts.

2. Executive Functioning Measures. All tests were administered per the guidelines presented in their respective administration manuals by trained research assistants. PennCNB tasks were part of a larger computerized battery that participants completed on a desktop with trained research assistant supervision.

***2.1 D-KEFS Subtest Measures*.** For TMT, the completion time in seconds for the fourth condition (letter-number switching), is thought to assess mental flexibility. For CWIT, completion time in seconds for fourth condition (inhibition/switching) is thought to assess inhibition and mental flexibility. For Tower, the total achievement score is thought to assess spatial planning and maintenance of instructional set. For DF, the total number of attempted designs is thought to assess initiation of problem solving behavior and fluency in generating visual patterns. For VF, the switching accuracy on the third condition (Fruits & Furniture) is

thought to measure mental flexibility. See Supplemental Table 1 for psychometric properties of tests.

***2.2 PennCNB N-Back Task*.** The n-back task is a working memory paradigm which also places demands on sustained attention and is commonly used in cognitive neuroscience.. Participants completed several rounds of the 0-back, 1-back, and 2-back tasks. Efficiency on the more demanding 2-back condition (i.e. a combination of both accuracy and speeded response) was included in analyses.

3. Imaging. DTI images were acquired using a 3.0 T Siemens Trio scanner with multiband echo planar imaging — voxel size = 2.0 × 2.0 × 2.0 mm, 64 slices, 137 direction, b-value = 1500 s/mm2, TE = 85 ms, TR = 2400 ms, FOV = 212 mm, and 12 non-diffusion volumes. Structural images (MPRAGE) were acquired during the same session – voxel size = 1.0 x 1.0 x 1.0 mm, 176 slices, TE = 2.52 ms, TR = 1900 ms, FOV = 250 mm.

3.1 GM Image Processing. MPRAGE data obtained from the NKI-RS dataset were available in their raw form. DICOM data were converted to the mgz format and MPRAGE images were automatically reconstructed in Freesurfer 5.3 (http://surfer.nmr.mgh.harvard.edu). Each subject was mapped into standard morphological space and cortical volume was generated for ROI analyses based on the Desikan-Killiany Atlas (Desikan et al., 2006). Cortical volumes were normalized for ICV, and averaged across hemispheres. Volumes from frontal and parietal regions were chosen using a priori hypotheses regarding their association with EF (Collette et al., 2005; Spreng, Sepulcre, Turner, Stevens, & Schacter, 2013; Neindam et al., 2012), and confirmed using measurement models for latent variables.

3.2 White Matter Image Processing. DTI obtained from the NKI-RS dataset have already been converted to NIfTI format and include the b-values and gradient directions. Because diffusion weighted images and echo planar imaging are more susceptible to the effects of eddy currents (Jehenson & Syrota, 1989; Jezzard, Barnett, & Pierpaoli, 1998), images were corrected for eddy currents using the FDT Diffusion Toolbox in FSL (Behrens et al., 2003; Behrens, Johansen-Berg, Jbabdi, Rushworth, & Woolrich, 2007) as part of TRACULA (Yendiki, et al., 2011). This corrected for motion between volumes and distortions.

For each subject, TRACULA registered reconstructed T1 images to diffusion DICOM data and generated automated probabilistic reconstruction of a set of standard major white-matter pathways using automatic segmentation statistics generated during reconstruction. Missing tracts were reinitialized and visually confirmed.

4. Structural Equation Modelling. Prior to confirmatory analysis, exploratory factor analyses on EF and GM measures were performed in SPSS Version 24. In order to verify that data were factorable, a maximum likelihood extraction with varimax rotation was used. For EF one measure from each subtest was selected for each iteration. Some subtests (i.e. CWIT, N-Back) had multiple scores that were appropriate for analyses. Each score was assessed separately, and then factor loadings were compared to determine which score fit the factor best. For the GM variable, all relevant regions from the frontal and parietal regions were selected (Neindam et al., 2012). The factor analytic adequacy of our data was evaluated by the KMO Measure of Sampling Adequacy (> .5) and the Bartlett's Test of Sphericity ($p < .001$). After determining that the factor had sufficient goodness-of-fit at or below statistical threshold ($p \leq .05$), the factor matrix was examined to confirm that all measures were loading at or above 0.40. For subtests that had multiple scores loading above .40, the one with the highest factor loading was chosen.

For the latent variable phase of the analyses AMOS Version 20 was utilized. A measurement model was used to assess the association between indicators and corresponding latent variables. Decisions regarding model fit were determined with respect to all measures, using commonly accepted cut-offs (Hu & Bentler, 1999), as follows. The following array of model fit indices and values were used to determine the final model: adjusted chi-square ($1 < Cmin/df < 5$), the comparative fit index ($CFI > .95$), and the root mean square error of approximation (RMSEA .02 - .08).

5. Pre-Analysis. Although the factor loading for the EF measure VF fell below the cut-off criteria (.382) during the exploratory phase, it was included in the confirmatory phase in the interest of thoroughness. When confirming the EF construct in SEM using AMOS software, it was determined that VF did not load appropriately (.294-.475) and was dropped from the construct.

When confirming the bilateral GM latent variable in AMOS, the model failed to converge with all indicators from the exploratory phase. Although the caudal anterior cingulate cortex was determined to be appropriately associated with the GM factor, it had the lowest factor loading (.564). When it was excluded from the latent variable, the model converged.

Supplementary Table 1. Psychometric properties of D-KEFS subtests.

	Test Retest *(r)*	Internal Consistency
TMT	0.38	0.60 - 0.81
Tower	0.44	0.56 - 0.78
CWIT	0.65	0.72 - 0.86
DF	0.32 – 0.58	--

Supplementary Table 2. EF indicator descriptive statistics.

	M(SD)	Skew	Kurtosis
TMT	.04(.95)	-2.12	5.51
Tower	.03(1.00)	-0.57	0.79
CWIT	.00(1.01)	-1.55	3.17
DF	.02(1.00)	0.69	1.35
N-Back	.01(.97)	-2.63	9.51

Supplementary Table 3. Group comparisons (younger vs. older) for EF indicators.

	Younger			Older		
	M(SD)	Skew	Kurtosis	M(SD)	Skew	Kurtosis
TMT	.20(.84)	-2.34	8.07	-.21(1.07)	-1.82	3.32
Tower	.10(.87)	-0.24	0.95	-.07(1.18)	-0.64	0.05
CWIT	.17(.86)	-1.39	3.13	-.27(1.17)	-1.42	1.95
DF	.13(1.08)	0.7	1.23	.17(.84)	0.25	0.01
N-Back	.05(1.01)	-2.76	10.07	-.05(.90)	-2.4	8.44