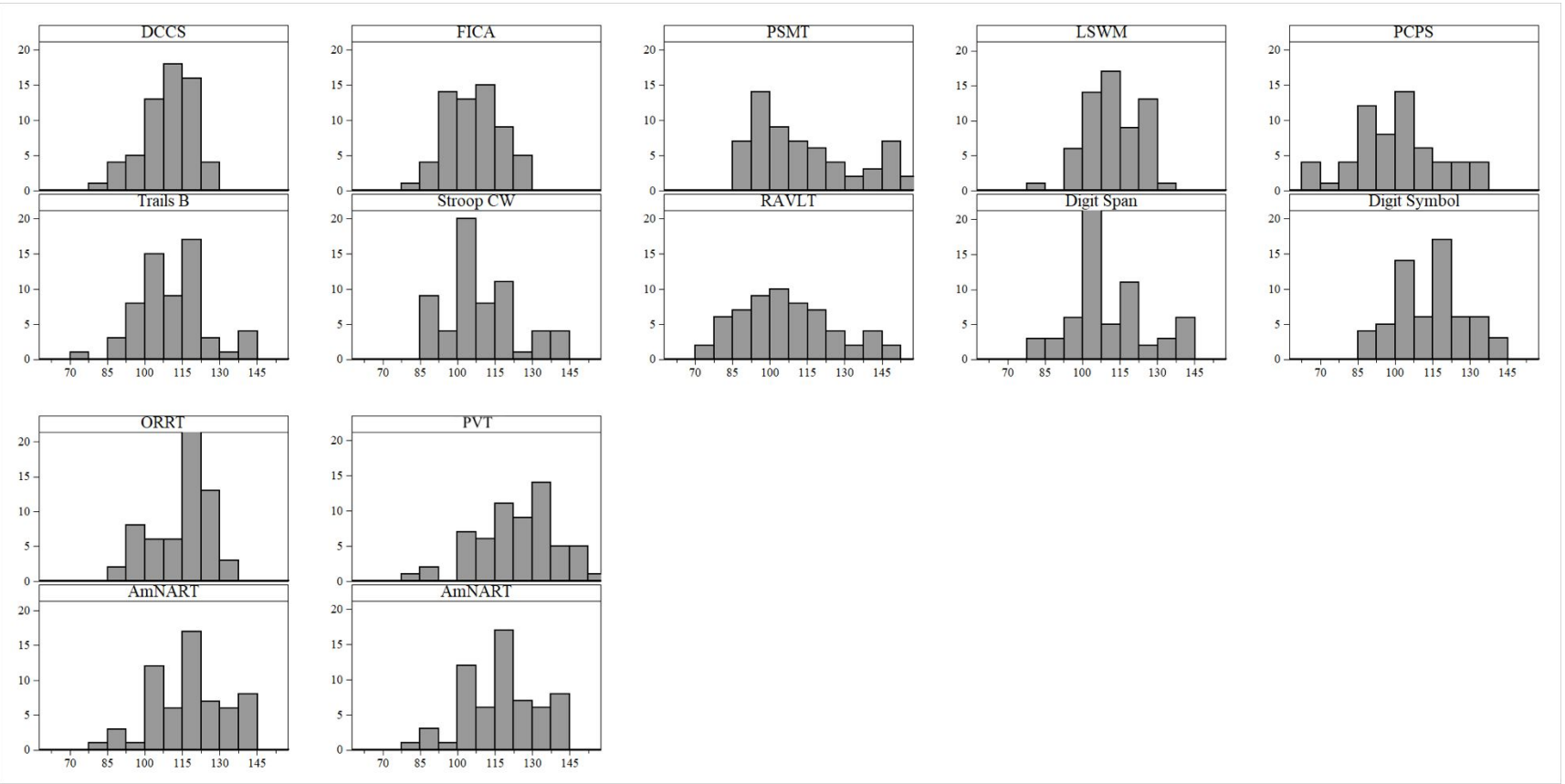
Supplementary Table 1. Descriptive statistics and Pearson's correlations among age-adjusted standard scores

	DCCS	FICA	PSMT	LSWM	PCPS	Trails B	Stroop C/W	RAVLT	Digit Span	Digit Symbol	PVT	ORRT	AmNART
Fluid Composite Measures													
DCCS	1.00												
FICA	.33**	1.00											
PSMT	.10	.00	1.00										
LSWM	.05	.22	.18	1.00									
PCPS	.43***	.39**	.33**	.32*	1.00								
Trails B	.22	.06	.14	.28*	.31*	1.00							
Stroop C/W	.31*	.10	.23	.22	.53***	.49***	1.00						
RAVLT	.37**	.27*	.48***	.32*	.23	.31*	.31*	1.00					
Digit Span	.09	.04	.03	.49**	.01	.37**	.15	.32*	1.00				
Digit Symbol	.21	.14	.37**	.17	.43***	.50***	.49***	.31*	.12	1.00			
Crystallized Composite Measures													
PVT	.24	.33**	.34**	.32*	.21	.19	.13	.52***	.28*	.25	1.00		
ORRT	.23	.28*	.31*	.29*	.19	.15	.13	.30*	.40**	.21	.70***	1.00	
AmNART	.18	.18	.36*	.14	.05	.04	.04	.34**	.31*	.21	.72***	.85***	1.00
Mean	108.79	106.39	112.89	111.70	99.93	108.77	107.95	106.54	108.59	112.97	123.46	114.66	115.49
SD	10.12	10.63	20.43	11.05	16.89	13.33	14.43	18.80	15.43	13.82	15.32	11.11	14.19
Min	79	83	86	80	65	76	85	70	80	85	84	87	80
Max	127	129	154	132	134	138	138	152	138	143	153	131	140

Notes. DCCS=Dimensional Change Card Sort Test; FICA=Flanker Inhibitory Control and Attention Test; PSMT=Picture Sequence Memory Test; LSWM=List Sorting Working Memory Test; PCPS=Pattern Comparison Processing Speed; PVT=Picture Vocabulary Test; ORRT=Oral Reading Recognition Test; AmNART=American National Adult Reading Test; Stroop C/W=Stroop Color-Word trial; RAVLT=Rey Auditory Verbal Learning Test (total learning trials 1-5); WMS-R=Wechsler Memory Scale–Revised; WAIS-III=Wechsler Adult Intelligence Scale-Third Edition. *$p < .05$; **$p < .01$; ***$p < .001$.

Boxed values are convergent validity correlation coefficients. All GS scores reflect appropriate age-adjusted normative scores that were converted to standard scores (M=100, SD=15).



Supplementary Figure 1. Histograms illustrating the distributions of age-adjusted scores on individual NIHTB-CB tests (top) paired with their corresponding GS (bottom) tests. Bins are defined by their boundaries, which correspond to 0.5 standard deviation units (e.g., 100-107.5, 107.5-115).

Legend: DCCS=Dimensional Change Card Sort Test; FICA=Flanker Inhibitory Control and Attention Test; PSMT=Picture Sequence Memory Test; LSWM=List Sorting Working Memory Test; PCPS=Pattern Comparison Processing Speed; ORRT=Oral Reading Recognition Test; PVT=Picture Vocabulary Test; Stroop CW=Stroop Color-Word trial; RAVLT=Rey Auditory Verbal Learning Test (total learning trials 1-5); AmNART=American National Adult Reading Test.

Supplementary Analyses

Methods

We carried out two sets of post hoc exploratory analyses to identify factors influencing the magnitude and direction of discrepancies between NIHTB-CB and GS Crystallized and Fluid Cognition Composite scores.

Subgroup analyses to test potential factors driving discrepancies between composite scores. We explored demographic characteristics that differed based on the size and direction of discrepancy of composite scores by stratifying the sample into three groups: NIHTB-CB = GS (discrepancy scores < 0.5 SD), NIHTB-CB < GS (≥ 0.5 SD), and NIHTB-CB > GS (≥ 0.5 SD). We calculated base rates for each discrepancy score group according to gender (i.e., male, female), race/ethnicity (i.e., white, black), age group (i.e., 60-69 years, 70-80 years), educational attainment (i.e., 12-16 years, 17-20 years). Gender and race/ethnicity differences were compared across discrepancy score groups using chi-square statistics. Age, educational attainment, and cerebral volumes were compared using ANOVA. These analyses are considered preliminary due to the small sample sizes and are provided for exploratory purposes only. Additionally, we compared the scores of individual subtests on the NIHTB-CB and GS (i.e., AmNART) Crystallized Cognition Composite using paired t-tests to determine whether a particular subtest was driving the discrepancy between composite scores.

Proportional differences between Fluid Cognition Composites. We additionally performed a non-parametric linear regression analysis (Passing & Bablok, 1983) to further investigate the positive trend in Fluid Cognition Composite discrepancy scores at higher values (NIHTB-CB > GS) that was observed in the Bland-Altman plot. Passing-Bablok regression enables differentiation between constant and proportional differences (i.e., larger discrepancies in particular ranges of scores) by estimating intercept and slope parameters with 95% confidence intervals. This method is advisable for method comparison studies because it allows for measurement error in both methods, including the reference test (Passing & Bablok, 1983).

Results

Subgroup analyses. Supplementary Table 2 shows demographic characteristics of participants in each discrepancy score group (NIHTB-CB < GS, NIHTB-CB = GS, and NIHTB-CB > GS). For the Crystallized Cognition Composite, only 3 participants obtained NIHTB-CB scores that were at least 0.5 SD below their GS scores (maximum NIHTB-CB < GS discrepancy = 10 points), so these participants were combined with the NIHTB-CB = GS group due to the small sample size. There were no significant differences between Crystallized Cognition Composite discrepancy score groups on any demographic variable, including gender ($\chi^2 = 0.28$, $p = .870$), race/ethnicity ($\chi^2 = 2.75$, $p = .253$), age (F [1, 59] = 0.06, $p = .802$), education (F [1, 59] = 0.29, $p = .595$), or cerebral volume (F [1, 53] = .317, $p = .545$ and F [1, 53] = .248, $p = .621$ for lateral ventricle and hippocampal occupancy, respectively). Examination of individual subtest scores revealed that the two oral reading tests (i.e., ORRT and AmNART) yielded equivalent means (mean difference = -0.84, 95% CI [-2.77, 1.09], SD = 7.53, $t = -0.87$, $p = .389$), but that the Picture Vocabulary Test differed significantly from AmNART (mean difference = 7.97, 95% CI [5.10, 10.83], SD = 11.18, $t = 5.57$, $p = .000$), suggesting that this test alone drives the discrepancy between NIHTB-CB and GS Crystallized Cognition Composite scores.

For the Fluid Cognition Composites, demographic comparisons indicate that age differed significantly across groups (F [2, 58] = 5.09, p = .009). Those with higher Fluid scores (>7.5 SD) on the NIHTB-CB tended to be younger (M = 64.92) than those with equal (M = 68.00) or higher GS scores (M = 70.55). Despite the very small sample size, racial differences also emerged across groups ($\chi^2 = 13.86$, $p = .001$), as five out of the six black participants (83%) performed significantly lower on the NIHTB-CB Fluid Cognition Composite (NIHTB < GS group). There were no significant group differences for gender ($\chi^2 = 0.03$, $p = .984$), years of education (F [2, 58] = 0.83, $p = .440$), or cerebral volumes (F [2, 27] = 4.232, $p = .025$ and F [2, 52] = 2.69, $p = .077$ for lateral ventricle and hippocampal occupancy, respectively).

Proportional differences between Fluid Cognition Composite scores. Supplementary Figure 2 shows the results of the Passing-Bablok regression analysis. The slope differed significantly from one

(1.54, 95% CI [1.13, 2.16]), confirming the presence of proportional differences between NIHTB-CB and GS Fluid Cognition Composite scores. The NIHTB-CB appears to overestimate performance for individuals with high scores and underestimate it among individuals with lower scores, although "lower" performance in this sample is still largely within the average range (minimum Fluid Cognition Composite scores: GS = 87 and NIHTB-CB = 77).

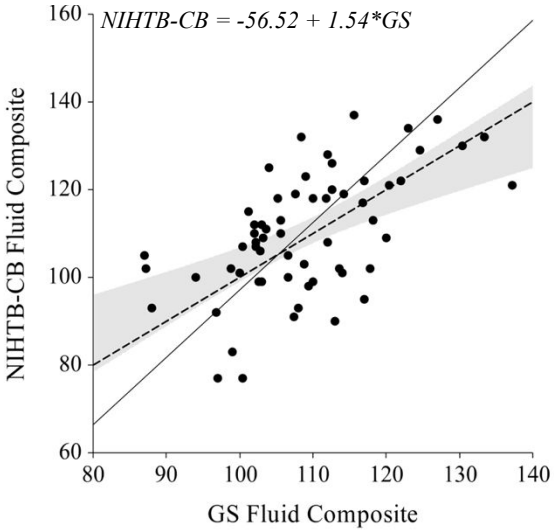
Supplementary Table 2. Demographic characteristics across discrepancy score groups

Crystallized Cognition Composite	NIHTB-CB < GS*		NIHTB-CB = GS (*n*=26)*		NIHTB-CB > GS (*n*=35)	
Gender, *n* (%)						
Female			18	45%	22	55%
Male			8	31%	13	37%
Race/ethnicity, *n* (%)						
White			24	44%	31	56%
Black			2	33%	4	67%
Education, M (SD)			16.58	(2.21)	16.23	(2.72)
12-16 years			12	40%	18	60%
17-20 years			14	45%	17	55%
Age, M (SD)			67.93	(5.56)	67.58	(5.10)
60-69			17	40%	25	60%
70-80			9	47%	10	53%
Lateral ventricles, M (SD)			2.01	(1.25)	1.84	(0.90)
Hippocampal occ., M (SD)			0.79	(0.07)	0.79	(0.06)

Fluid Cognition Composite	NIHTB-CB < GS (*n*=14)		NIHTB-CB = GS (*n*=30)		NIHTB-CB > GS (*n*=17)	
Gender, *n* (%)						
Female	9	23%	20	50%	11	28%
Male	5	24%	10	48%	6	29%
Race/ethnicity, *n* (%)						
White	9	16%	29	53%	17	31%
Black	5	83%	1	17%	0	0%
Education, M (SD)	15.71	(2.64)	16.40	(2.61)	16.88	(2.20)
12-16 years	8	27%	15	50%	7	23%
17-20 years	6	19%	15	48%	10	32%
Age, M (SD)	70.55	(4.99)	68.00	(5.46)	64.92	(3.75)
60-69	7	17%	21	50%	14	33%
70-80	7	37%	9	47%	3	16%
Lateral ventricles, M (SD)	2.08	(1.12)	2.15	(1.28)	1.44	(0.50)
Hippocampal occ., M (SD)	0.79	(0.07)	0.76	(0.07)	0.82	(0.04)

Percentages are rounded to the nearest whole number. ANOVA and chi-square statistics were calculated for continuous variables and categorical variables, respectively. Welch's ANOVA was used for analyses with lateral ventricle volumes. All other continuous variables had equal variances.

*For Crystallized discrepancy scores, individuals in the NIHTB-CB < GS group (*n* = 3, maximum discrepancy = 10) were combined into the NIHTB-CB = GS group due to the small sample size.



Supplementary Figure 2. Results of Passing-Bablok regression indicate significant proportional differences between NIHTB-CB and GS Fluid Cognition Composite scores. The regression line and 95% confidence intervals are shown as a solid line and shaded regions. The line of identity is shown as a dotted line. Regression coefficients (95% CI): intercept -56.52 (-123.39, -9.83) and slope 1.54 (1.13, 2.16).