Overnight retest-learning predicts long-term cognitive trajectory in older adults

David L. Woods[1], Kathleen Hall[1], Isabella Jaramillo[1], Mike Blank[1], Kristin Geraci[1], Peter Pebler[1], and David K. Johnson[2]

[1]Neurobehavioral Systems, Inc., Berkeley, CA;

[2]UC Davis ADRC, Walnut Creek, CA

Supported by NIA R44AG097322. Correspondence: drdlwoods@neurobs.com

Key Findings

• We examined the cognitive trajectories of 172 normal older adults on 15 subtests of California Cognitive Assessment Battery (CCAB) administered at intervals of one day, 6 months, 18 months, and 30 months.

• Nearly all subjects (98.3%) improved their test scores on overnight retest with a strong correlation (r = 0.61) between overnight retest gain and long-term performance at 6-30 months. Participants were divided into quartiles based on the magnitude of overnight retest learning. Q2-Q4 participants showed large learning effects (up to +0.91 z) and retained 40–50% of overnight gains over the 30-month follow-up. In contrast, Q1 participants fell below baseline at longer retest intervals despite significant overnight retest gain (+0.21 z, $t = 10.03$, $p < 10^{-10}$).

• Significant learning effects were seen in all five cognitive domains (Executive Function, Processing Speed, Episodic Memory, Language, and Speech Fluency). Retest gains were domain-specific: they predicting improved long-term performance in that domain but were uncorrelated with improvements in other domains.

• These findings support overnight retesting as a low-burden prognostic tool for trial enrichment. After adjustment for baseline cognition and demographics, a 1-SD decrease in omnibus overnight retest-learning was associated with a 2.7-fold increase in the odds of bottom-quartile 30-month cognitive trajectory (OR = 2.70, 95% CI 1.65–4.42; AUC = 0.82), a predictive value (r ~ 0.61) which is similar to the combined predictive value of p-tau217 and PET imaging (Ossenkoppele et al., 2025).

Abstract

Background. Short-term retest learning on neuropsychological tests, improvements observed when subjects retake the same tests within days, have been proposed as biomarkers of cognitive integrity. Reduced retest learning has been associated with accelerated cognitive decline and elevated Alzheimer's Disease (AD) risk markers including amyloid and tau deposition. Most prior work has examined retest learning in memory-focused tests at one week intervals in samples enriched for mild cognitive impairment. Here we investigated whether overnight retest learning in a comprehensive, 2-hr assessment would predict cognitive trajectories in cognitively unimpaired older adults over a 2.5-year follow-up.

Methods. 172 cognitively unimpaired older adults completed 14 tests in the California Cognitive Assessment Battery (CCAB) at baseline, overnight retest, and at 6, 18, and 30 months. An omnibus composite score was constructed as the mean of z-scores in five cognitive domains: Executive Function (EF), Language/Story Memory (LS), Episodic Memory (EM), Processing Speed (PS), and Speech Fluency (SF). Trajectories were examined as (a) raw change from baseline and (b) baseline-conditioned residuals using standardized regression-based change. Regression-to-the-mean (RTM) contributions were quantified by comparing observed slopes of baseline-vs-change relationships to slopes predicted under the no-learning null.

Results. The full cohort showed substantial 1-day retest learning on the omnibus composite (+0.55 z, $t(171) = 25.12$, $p < 10^{-62}$). Subjects were stratified into quartiles (Q1-Q4) based on overnight learning magnitude. Different quartiles did not differ in enrollment performance (ANOVA $F(3,168) = 1.02$, $p = 0.39$), but diverged thereafter, with the magnitude of the overnight learning correlating strongly ($r = 0.61$) with longitudinal trajectory. Retest effects partially decayed and stabilized at 6 to 30 months in higher quartiles Q2-Q4. In contrast, Q1 participants, with the smallest overnight retest gain (+0.21 z, $t = 10.03$, $p < 10^{-10}$), fell below baseline by 6 months and remained there for the next two years. Regression to the mean (RTM) analysis indicated that the majority of retest gain reflected domain-specific learning. Each domain's learning effect predicted its own 30-month trajectory (within-domain $r = 0.56$) independently from other domains' trajectories (cross-domain $r = 0.05$). After adjustment for baseline cognition and demographics a 1-SD decrease in omnibus overnight retest-learning magnitude was associated with a 2.7-fold increase in the odds of falling into bottom-quartile (Q1) trajectory (OR = 2.70, 95% CI 1.65–4.42; AUC = 0.82).

Conclusions. The magnitude of overnight retest learning on the CCAB battery predicted cognitive outcomes at 6 to 30 months in cognitively unimpaired older adults with a predictive accuracy approaching that of multi-modal AD biomarker panels.

Introduction

Retest learning on neuropsychological tests, improvements in performance when the same tests are administered repeatedly, reflects the retention and consolidation of content and procedures experienced in initial test exposure (Tort-Merino et al., 2024). At retest intervals of ~6 months to 4.5 years, the average retest learning gain across cognitive measures is approximately 0.25 standard deviations, with substantial heterogeneity by cognitive domain (Calamia et al., 2012; Scharfen et al., 2018). Duff and colleagues have found that the magnitude of tretest learning gains at 1-week intervals differentiate cognitively intact older adults from those with mild cognitive impairment (MCI) and Alzheimer's disease (AD), correlate with amyloid deposition on PET imaging, and predict subsequent cognitive decline (Duff et al., 2014, 2017; Hassenstab et al., 2015; Hammers et al., 2021).

Several methods for quantifying retest learning have been advanced, including simple discrepancy scores, reliable change indices (RCI; Jacobson & Truax, 1991), and standardized regression-based change indices that correct follow-up scores for baseline performance and demographics (Chelune et al., 1993; Crawford & Garthwaite, 2007). Each approach involves tradeoffs: raw change scores are subject to contamination from regression-to-the-mean (RTM), while regression-based residuals risk absorbing a portion of the learning signal they are designed to detect. Duff and colleagues (2019) compared seven change formulae against neuroimaging biomarkers and found that regression-based analyses showed the strongest relationships to amyloid deposition and hippocampal volume.

The clinical utility of retest learning depends on whether it predicts cognitive trajectories over the multi-year timescales relevant to early intervention. Here, we address this question using data from the California Cognitive Assessment Battery (CCAB; Woods et al., 2024) in 172 cognitively unimpaired older adults assessed at baseline, overnight retest, and follow-up waves at 6, 18, and 30 months. Unlike previous studies of retest learning focused on memory measures, the CCAB included 15 tests to examine retest learning in five different cognitive domains identified with confirmatory factor analysis (Woods et al., 2026b). We adopt a comprehensive analytic strategy that first presents results in raw change scores — minimizing residualization-related assumptions — and confirms findings using classical baseline-conditioned residuals as a supporting analysis. We then explicitly quantify the contribution of regression-to-the-mean to the observed signal using a classical-test-theory framework and evaluate the utility of overnight retest learning in identifying participants at risk of subsequent cognitive decline.

Methods

Participants

Participants were 172 community-dwelling older adults who were early enrollers in the CCAB longitudinal normative study who had completed five assessment waves: baseline, overnight retest, and 6, 18, and 30 month follow up sessions. All participants were classified as neurologically normal and cognitively unimpaired at enrollment based on standard clinical criteria. All participants provided written informed consent.

The study included two cohorts: 90 military veterans recruited through the Veterans Administration (VA) and 82 community-dwelling volunteers recruited by Neurobehavioral Systems (NBS Normative, see Table 1).The two subsamples differed significantly in age ($t = 3.32$, $p = .001$), sex composition ($\chi^2 = 53.3$, $p < 10^{-13}$), education ($t = -2.82$, $p = .005$), vocabulary ($t = -5.73$, $p < 10^{-6}$) and race, motivating the inclusion of research group as a covariate in later demographic-adjustment analyses (see below).

Table 1. Demographic characteristics of the study cohort (N = 172 cognitively unimpaired older adults) and the two contributing subsamples.

Characteristic	Full cohort (n = 172)	VA (n = 90)	NBS Normative (n = 82)
Age, years — mean (SD)	70.8 (6.4)	72.4 (6.6)	69.2 (5.9)
range	56–89	61–89	56–84
Education, years† — mean (SD)	17.0 (2.5)	16.5 (2.6)	17.6 (2.3)
Vocabulary (QSIN) — mean (SD)	42.3 (7.0)	39.7 (6.5)	45.1 (6.5)
Sex — n (%)			

Female	77 (44.8%)	16 (17.8%)	61 (74.4%)
Male	95 (55.2%)	74 (82.2%)	21 (25.6%)
Race/ethnicity — n (%)			
White	129 (75.0%)	54 (60.0%)	75 (91.5%)
Black	16 (9.3%)	14 (15.6%)	2 (2.4%)
Asian	16 (9.3%)	15 (16.7%)	1 (1.2%)
Other	11 (6.4%)	7 (7.8%)	4 (4.9%)
Hispanic/Latino	9 (5.2%)	8 (8.9%)	1 (1.2%)

CCAB Domain and Omnibus Composites

The CCAB is a proctored, telemedical test battery (Woods et al., 2024) that was administered in participants' homes or at a VA laboratory. Assessments included a selection of 15 subtests generating 62 measures. Domain z-scores, based on 7 to 16 measures per domain, were computed for Executive Function (EF), Language/Story Memory (LS), Memory (EM), Processing Speed (PS), and Spoken Fluency (SF) based on a five-domain structure supported by bifactor confirmatory factor analysis on the entire 1,916 participant CCAB normative sample (Woods et al., 2026b). The omnibus (omnibus) composite was constructed as the unweighted mean of the five domain z-scores at each wave.

Quartile Stratification

Subjects were stratified into quartiles (Q1-Q4) ordered by the magnitude of overnight retest learning gain in the omnibus score. We stratified participants by change scores rather than residualized retest-learning trajectories as the primary stratification variable for two reasons: (a) raw change makes no residualization assumptions, providing the most transparent characterization of cohort performance; and (b) raw-change quartiles showed statistically equivalent baselines scores across groups (ANOVA $F(3,168) = 1.02$, $p = 0.39$), allowing trajectory comparisons unconfounded by baseline imbalance.

Trajectory Metrics

Two complementary trajectory metrics were computed at each wave. Raw change scores were defined per subject as (wave z − enrollment z), with no residualization. Demographically corrected baseline-conditioned residuals were defined as the residuals from cohort-level OLS regressions of each wave's z-score at enrollment (Chelune et al., 1993) additionally corrected for age, education, gender, race, vocabulary, and research group.

Statistical Analyses

Cohort-mean change scores were tested against zero using one-sample t-tests. Quartile differences in mean follow-up trajectories were tested using one-way ANOVA across all four quartiles. Effect sizes are reported as Cohen's d. The correlation of overnight learning to mean-follow-up-change was tested against a “no-learning null distribution” generated by 10,000 random permutations of subject retest learning changes (fixed seed 20260701 for reproducibility). To express the prognostic relationship in odds-ratio form we fit continuous logistic regressions in where the outcome was the binary indicator of bottom-quartile (Q1) mean 6-30 month trajectory and the predictor was domain-specific overnight retest gain. Following biomarker conventions, odds ratios are reported per 1-SD decrease in overnight gain so that $OR > 1$ corresponds to elevated decline risk. Adjusted models corrected for baseline performance and the influence of age, gender, education, vocabulary, race (Black, Asian, Other; White = reference), and research group (VA vs. NBS Normative). Discrimination is reported as area under the receiver operating characteristic curve (AUC). Analyses were conducted in R (v4.3) and Python (v3.12).

Results

Overnight retest learning effects

As shown in Figure 1 participants showed substantial overnight practice effects on the omnibus composite (black line, +0.55 z, $t(171) = 25.12$, $p < 10^{-62}$) that decayed substantially by 6 months and stabilized for the remaining 24 months. All four follow-up means differed significantly from zero, indicating a persistent cohort-level elevation above baseline performance lasting 30 months. Domain-specific patterns differed substantially. Memory (red line) showed the largest 1-day practice effect (+0.99 z, t = 26.3) and the most preserved long-term elevation (30 months = +0.60 z, t = 12.3). Language/Story Memory and Processing Speed showed large overnight effects with more modest retention, while Speech Fluency and Executive Function showed smaller initial effects that largely decayed by 6 months.

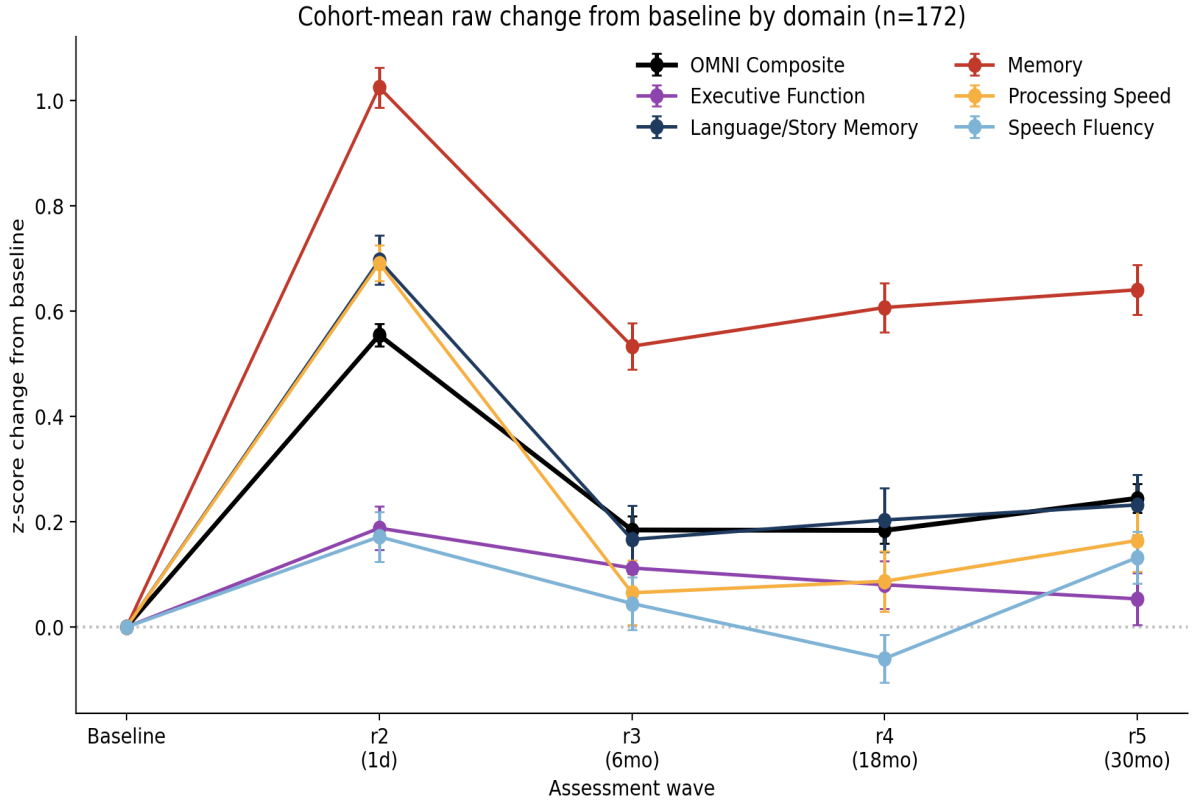

Figure 1. Participant raw change from baseline by domain (n = 172). The omnibus composite shows a +0.55 z mean gain at the 1-day retest (r2) that decayed to a stable +0.17–0.21 z plateau from six to 30 months. Memory showed the largest and most persistent practice effects; Executive Function shows the smallest, returning to baseline by 30 months. Error bars show ±1 SEM.

Stratification by 1-day retest-learning: omnibus trajectories

Figure 2A shows omnibus scores for the four quartiles of participants, stratified by the magnitude of overnight learning. The quartiles did not differ significantly in baseline performance (ANOVA on r1: $F(3,168) = 1.02$, $p = 0.39$) but, by construction, diverged sharply in the magnitude of overnight learning. Shown relative to baseline in Figure 2B, Q1 gained +0.19 z, Q2 +0.46 z, Q3 +0.64 z, and Q4 +0.91 z. Between six and 30 months Q2-Q4 regressed toward the cohort mean, with rank-ordering preserved at subsequent waves. However, Q1 fell below baseline at 6-months without subsequent recovery.

The cohort-wide correlation between overnight gain and mean follow-up change was r = 0.61, $t(170) = 10.14$, $p < 10^{-16}$). Under a no-learning null generated by 10,000 random permutations of subject overnight gain, the observed correlation fell 8.1 standard deviations above the null distribution mean.

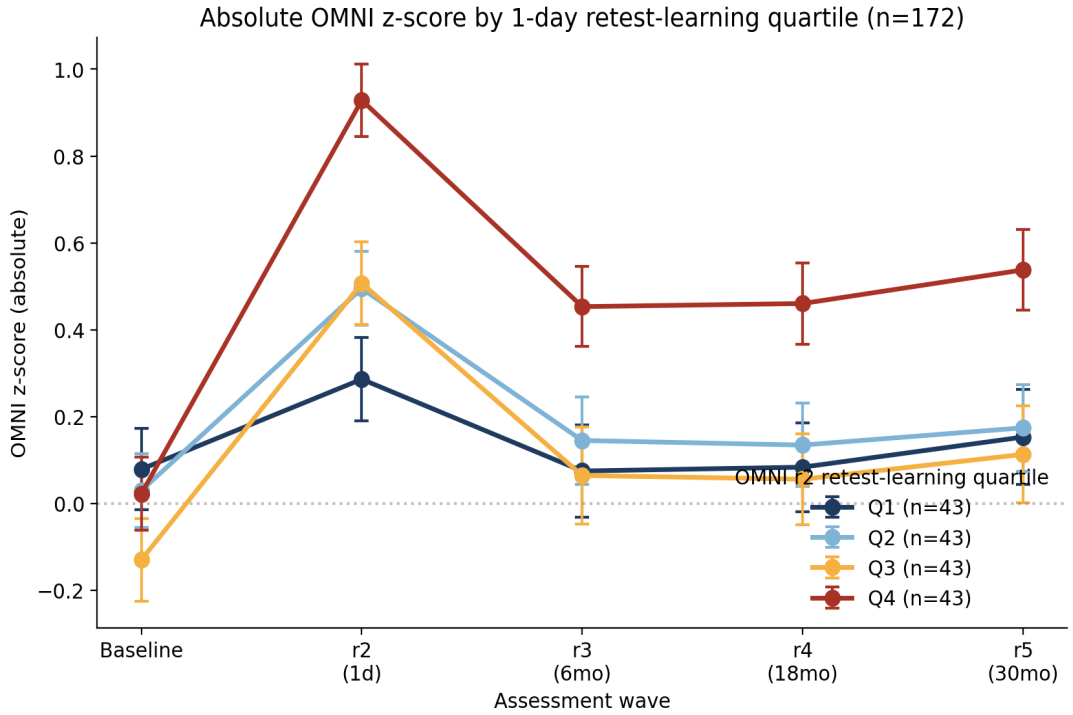

Figure 2A. Absolute omnibus z-score by 1-day retest-learning quartile (n = 172), showing each quartile's actual baseline mean. At 6 months the quartiles diverge dramatically by construction with rank ordering preserved from six to 30 months. Error bars show ±1 SEM.

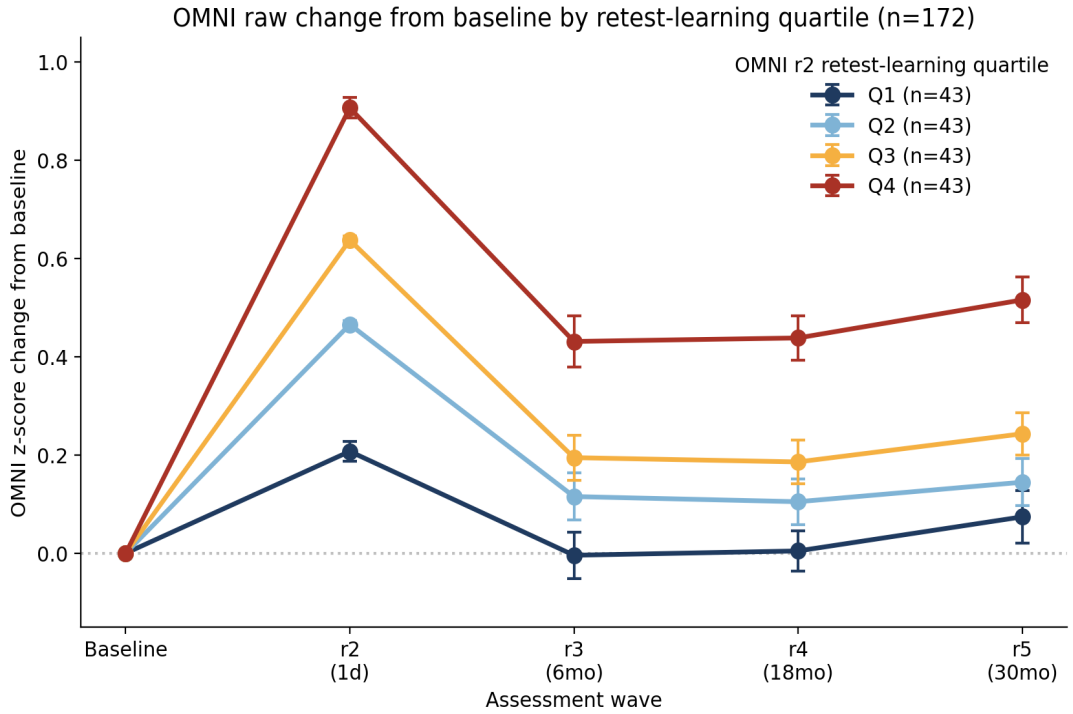

Figure 2B. omnibus raw change from baseline. Q4 subjects retain approximately half their initial 1-day gain through 30 months (+0.52 z at r5). Q1 subjects return to baseline by 6 months and remain at baseline for the next 24 months. The 30-month Q4-vs-Q1 gap was 0.44 z (Cohen's d = 1.34); the gap in mean 6–30 month trajectory was 0.44 z (d = 1.67). Error bars show ±1 SEM.

Domain trajectories

What cognitive domains were responsible for the improvement in omnibus scores? Figure 3A shows domain-by-domain trajectories stratified by omnibus retest-learning quartile (not domain quartile) and reveals heterogeneous trajectories across the five cognitive domains, while Figure 3B shows corresponding raw changes from baseline. At baseline the four quartiles clustered tightly within each domain — consistent

with the equivalent-baselines property documented for the omnibus composite — with the exception of Language/Story Memory, where the Q4 baseline was significantly below cohort mean and Q1 significantly above (i.e., subjects with lowest LS enrollment scores showed the largest retest gains).

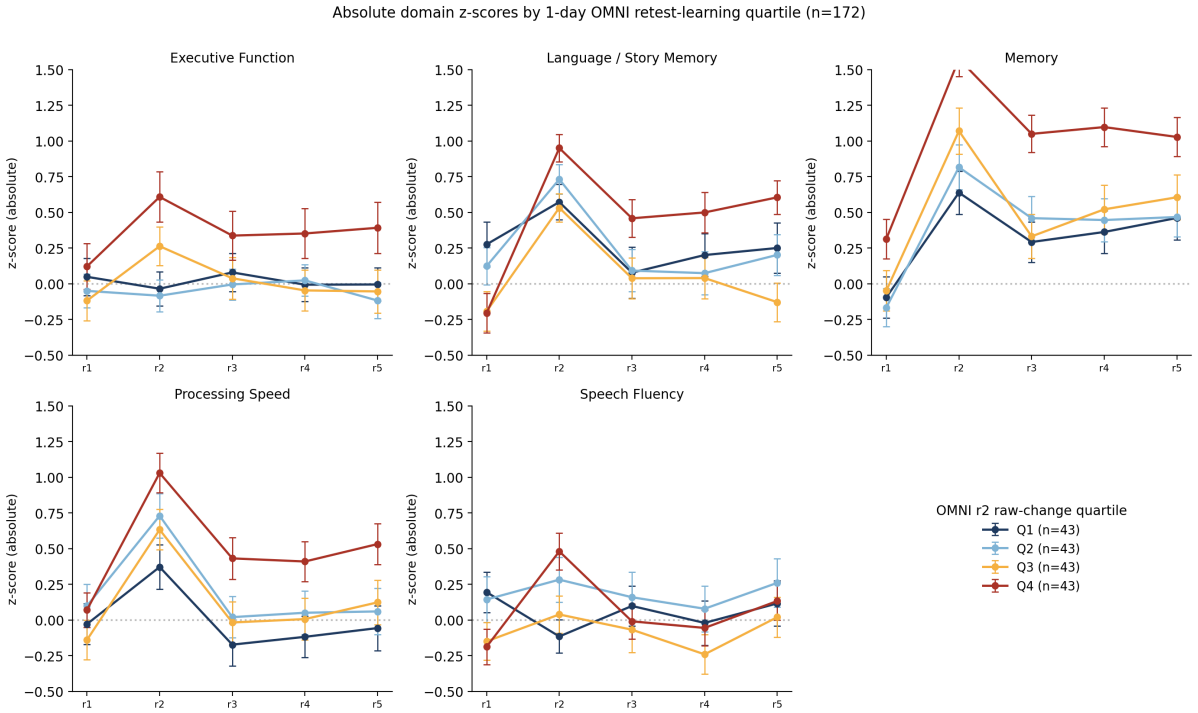

Figure 3A. Absolute domain z-scores at each wave by 1-day omnibus retest-learning quartile (n = 172). No baseline correction. Memory showed the most dramatic and most stable Q4 separation, persisting at near-full magnitude through 30 months. Processing Speed showed a large initial separation that retains substantial differentiation at 30 months. Language/Story Memory shows visible baseline imbalance (Q4 low, Q1 high at baseline) reflecting floor-recovery selection. Executive Function shows the noisiest, least-differentiated trajectories. Error bars show ±1 SEM.

Figure 3B shows retest gains for the different quartiles. All domains, except Executive Function, showed large overnight learning effects that persisted in subsequent test waves. The domains that differentiated the quartiles were different than those showing the largest absolute gain: at 30 months, Q4-vs-Q1 interquartile differences ranged from 0.16 z (Memory) to 0.84 z (Language/Story Memory).

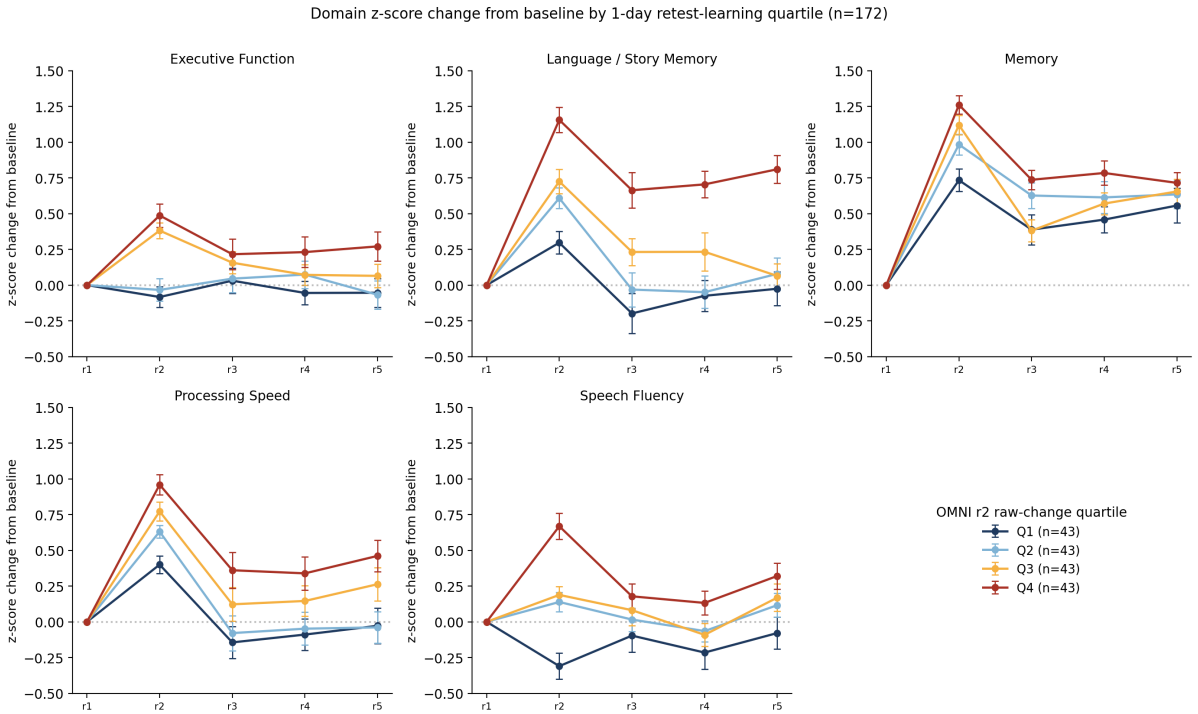

Figure 3B. Domain z-score change from baseline by 1-day retest-learning quartile (n = 172), based on the omnibus composite overnight learning effect magnitude. Memory showed the largest gains overall with modest quartile

differentiation. Language/Story Memory and Processing Speed show the strongest persistent quartile separations at 30 months, with Executive Function showing the weakest persistent quartile differentiation. Error bars show ±1 SEM.

Regression-to-the-mean

To quantify the contribution of RTM to the observed overnight learning effects, we compared observed learning slopes regressed on baseline scores against the slopes predicted under a classical-test-theory no-learning null (slope = TRR − 1) for each participant. Results are summarized in Table 2 along with test-retest reliability measures (enrollment vs. retest). For the omnibus composite, only a modest fraction (15%) of the retest learning effect could be attributed to RTM, with the remainder reflecting genuine subject-specific learning although the Pearson correlation between baseline and overnight learning scores statistically indistinguishable from zero (r = −0.11 (p = 0.15).

RTM contributions varied substantially in different cognitive domains. Memory showed essentially no RTM contribution (3%) with a near-zero baseline-change correlation (r = −0.01) and large overall practice effects (+0.99 z). Processing Speed showed modest RTM contribution (26%). At the other extreme, Language/Story Memory showed an observed slope steeper than the pure-RTM prediction (% RTM = 127%). This pattern was not driven by ceiling truncation: only 2% of LS subjects scored above z = 1.5 on enrollment, but reflected floor-recovery dynamics: subjects in the bottom LS quartile gained +1.28 z on overnight retest vs. +0.29 z for high-baseline subjects. At baseline, performance on the LS tasks for some subjects was characterized by hesitant speech with frequent errors, consistent with first-exposure encoding and retrieval failures on novel story material, that were corrected on retest. Mechanistically, this is distinct from classical RTM, which assumes a stable underlying true score with random measurement error; floor-recovery instead reflects state-dependent variance in initial measurement that is partially resolved by retest. The two domains with the cleanest retest-learning signals — Memory and Processing Speed — also showed the largest positive net slopes (+0.114 and +0.079), indicating that even high-baseline subjects gained more at overnight retest than RTM alone would predict, consistent with a learning advantage for more capable subjects.

Table 2. *Test-retest reliability and regression-to-the-mean diagnostics for the omnibus composite and five cognitive domains, n = 172.*

Domain	TRR	Mean Δr2	SD(Δr2)	Obs slope	RTM-pred slope	Net slope	% RTM[1]	r(r1,Δr2)	Q4−Q1 r5 Δ[2]
omnibus Composite	0.90	+0.55	0.28	−0.032	−0.099	+0.068	32%	−0.07	+0.44
Executive Function	0.83	+0.19	0.54	−0.136	−0.171	+0.035	79%	−0.23	+0.32
Language/Story Memory	0.76	+0.70	0.61	−0.423	−0.241	−0.182	175%	−0.65	+0.84
Memory	0.88	+1.02	0.50	−0.024	−0.125	+0.101	19%	−0.04	+0.16
Processing Speed	0.89	+0.69	0.45	−0.026	−0.108	+0.083	24%	−0.05	+0.49
Speech Fluency	0.77	+0.17	0.62	−0.249	−0.231	−0.019	108%	−0.37	+0.40

[1] *% RTM = Observed slope / RTM-predicted slope × 100. Values represent the proportion of the baseline vs. overnight retest relationship attributable to classical regression-to-the-mean under a no-learning null model.* [2] *Q4−Q1 r5 Δ: difference in mean difference at the 30-month wave between top and bottom omnibus r2 raw-change quartiles.*

Demographic effects on retest-learning

Table 4 shows demographic effects on unregressed change scores. Demographic influences were modest, jointly explaining only 3-15% of variance in 1-day change and 6-40% of variance in mean follow-up change for most domains. The three main findings were (1) increased age predicted smaller Memory practice effects; (2) higher vocabulary predicted better long-term omnibus trajectory; and (3) the older VA subsample showed steeper Processing Speed decline. By comparison, demographic factors explained ~30–50% of variance in absolute baseline cognitive scores in these cohort, i.e, 5-10 times larger than demographic effects on change scores.

Table 4. *Demographic effects on omnibus and domain change scores in 172 cognitively unimpaired older adults, adjusting for research-group composition. Each row is a single multiple regression of the indicated*

*change score on age, vocabulary, female sex, race (Black, Asian, Other; white = reference), and research group (VA veteran subsample, n = 90; NBS Normative community subsample, n = 82; NBS = reference). Standardized β coefficients shown. Panel A: 1-day retest change. Panel B: Mean change at the 6-, 18-, and 30-month follow-up waves. R^2 values give the cumulative variance explained by all seven covariates jointly. * p<.05, ** p<.01, *** p<.001.*

Panel A. Demographic effects on 1-day (r2) change

Domain	R^2	β Age	β Vocab	β Female	β Black	β Asian	β Other	β VA
omnibus	0.07	−0.14	+0.10	−0.06	−0.05	+0.11	−0.03	−0.16
EF	0.06	+0.04	−0.02	−0.07	+0.13	−0.03	+0.20*	−0.05
LS	0.10*	+0.00	−0.02	−0.10	−0.14	+0.19*	−0.18*	−0.08
Memory	0.10*	−0.24**	+0.08	+0.02	−0.11	−0.01	+0.03	−0.05
PS	0.14***	−0.12	+0.21*	−0.06	−0.01	−0.00	−0.02	−0.22*
SF	0.02	−0.08	+0.04	+0.09	+0.03	+0.06	−0.04	−0.00

Panel B. Demographic effects on mean follow-up (r3+r4+r5)/3 change

Domain	R^2	β Age	β Vocab	β Female	β Black	β Asian	β Other	β VA
omnibus	0.16***	−0.09	+0.20*	+0.10	+0.12	+0.06	−0.12	−0.20*
EF	0.05	+0.05	+0.07	+0.01	+0.13	−0.08	+0.03	+0.15
LS	0.10*	−0.00	+0.19*	−0.00	−0.02	+0.06	−0.24**	+0.01
Memory	0.07	−0.21**	−0.06	+0.09	−0.00	−0.06	+0.05	+0.01
PS	0.39***	−0.05	+0.07	+0.05	+0.08	+0.08	+0.01	−0.59***
SF	0.06	−0.03	+0.20*	+0.11	+0.12	+0.12	−0.09	+0.07

Note. The two subsamples differ substantially in sex composition (VA = 82% male; NBS Normative = 74% female), in recruitment source (veterans vs. community volunteers), and in long-term cognitive trajectories (β_VA = −0.20 for omnibus follow-up change, p<.05; β_VA = −0.59 for PS follow-up change, p<.001).

However, the VA subsample showed reliably smaller univariate retest learning than the NBS sample (omnibus Δr2 = +0.49 vs +0.61 z, p = 0.006), driven primarily by Processing Speed (d = −0.71). This subsample difference was largely mediated by age, vocabulary, and education, which distinguished the VA and NBS cohorts, but a residual VA-specific Processing Speed deficit persisted under full demographic adjustment (β_VA = −0.24, p = 0.017). All inferential analyses therefore included research group as a covariate to ensure that domain-level prognostic estimates were not confounded by subsample composition.

Trajectories under baseline-conditioned and demographic residualization

To confirm that the trajectories patterns observed in raw change scores held under classical psychometric residualization, we re-computed quartile trajectories using baseline-conditioned residuals at each follow-up wave (Chelune et al., 1993) with demographic adjustment for the predictors in Table 4. Quartile assignments remain based on omnibus overnight learning magnitude to preserve equivalence with previous analyses. Figure 4 shows the result for the omnibus composite. The Q4-vs-Q1 gap at r5 was 0.38 z under baseline-conditioned and demographic residualization (versus 0.44 z in raw change scores), with Q4 stabilizing at +0.24 z and Q1 at −0.14 z. The smaller absolute magnitude reflects residualization removal of between-subject variance attributable to baseline and demographic factors; the qualitative pattern of persistent quartile separation across all follow-up waves is similar to that of the raw-change analyses.

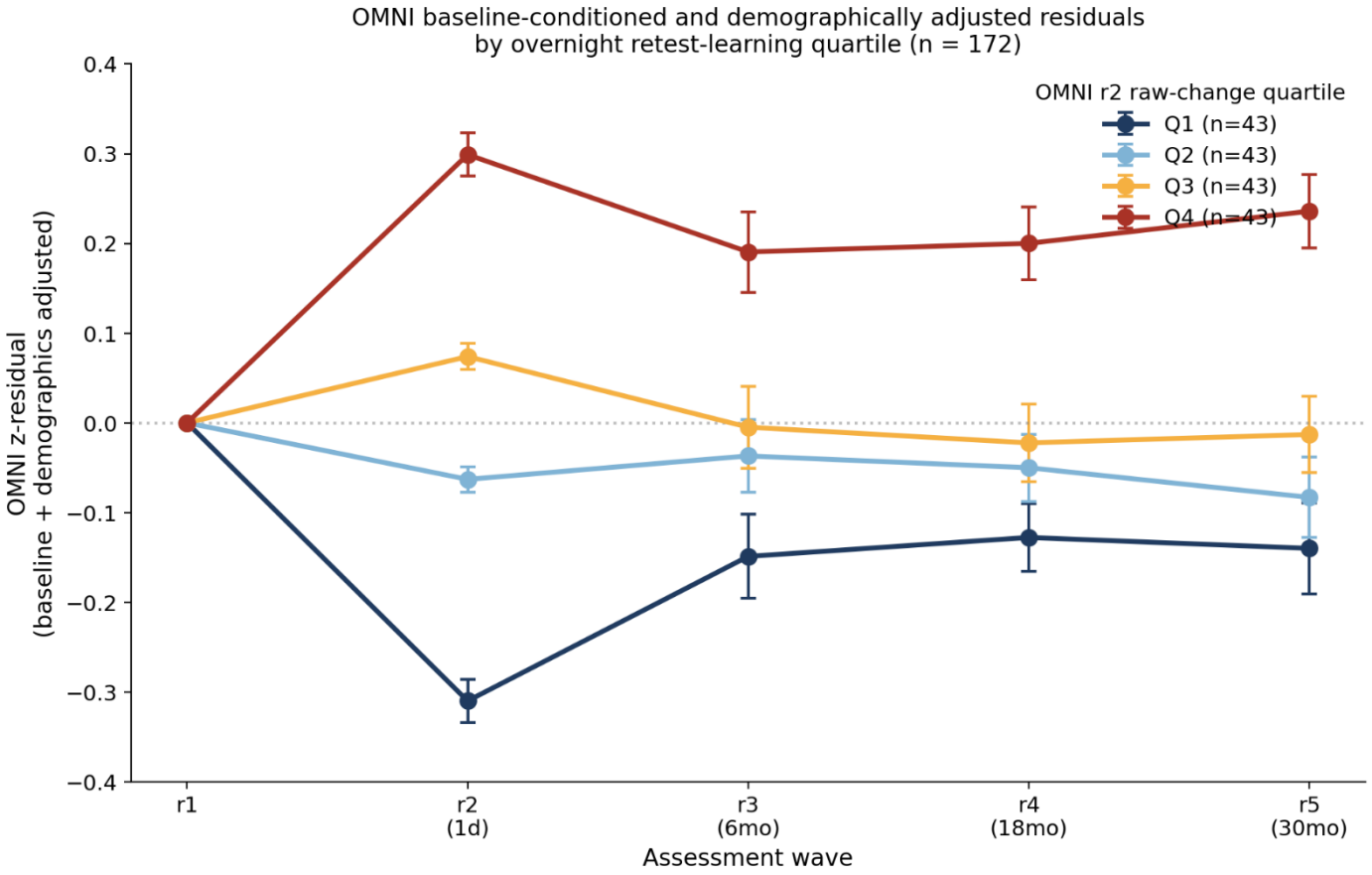

Figure 4. Baseline-conditioned and demographically adjusted change residuals by raw-change quartile (n = 172). Quartile assignments are based on raw omnibus overnight learning magnitude, i.e., the same quartile membership as in Figures 2 and 3. The Q4-vs-Q1 gap of 0.38 z at 30 months was smaller in absolute magnitude than with raw-change analysis (Figure 2B; gap = 0.44 z) because demographic adjustment removed trajectory variance attributable to age, vocabulary, and cohort-composition differences. The qualitative pattern of persistent quartile separation across all follow-up waves was preserved under this conservative adjustment, indicating that the prognostic signal is not attributable to either RTM or demographic confounding. Error bars show ±1 SEM.

Figure 5 shows domain baseline-conditioned and demographically adjusted residual trajectories which preserved the qualitative pattern of persistent Q4-vs-Q1 separation seen in Figure 3B. The residual Q4-vs-Q1 30-month gaps narrowed across domains: Language/Story Memory showed the largest residual gap (0.59 z), with Executive Function, Processing Speed, and Speech Fluency at 0.28–0.30 z and Memory lowest (0.22 z). Processing Speed showed substantial attenuation. (adjusted 30-month gap = 0.30 z vs. 0.49 z raw), reflecting the disproportionate contribution of the older VA cohort to long-term Processing Speed decline, which was partialed out by the demographic adjustment.

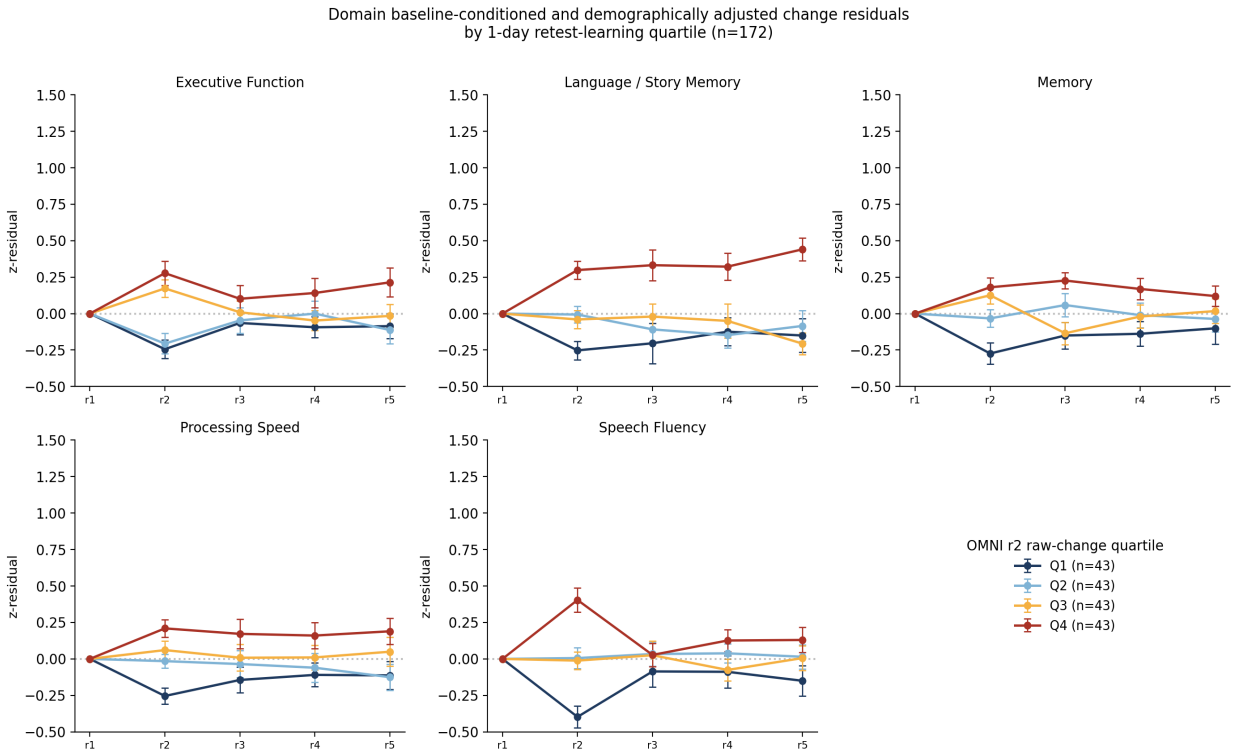

Figure 5. Domain baseline-conditioned and demographically adjusted change residuals by overnight learning quartile (as Figures 2 and 3). At each follow-up wave, the domain-specific raw change score was residualized on the corresponding domain baseline plus demographic covariates (age, vocabulary, female sex, race, research group), removing both baseline-driven RTM and demographically-predicted variance. The five domains showed similar magnitudes of persistent residual quartile differentiation: Language/Story Memory and Executive Function showed the largest 30-month Q4–Q1 residual gaps (0.59 z); Executive Function, Processing Speed, and Speech Fluency clustered at 0.28–0.30 z, with Memory lowest (0.22 z). Error bars show ±1 SEM.

Specificity of retest-learning

Figure 6 shows the matrix of correlations between overnight learning in each domain and its mean 6–30 month trajectory. Under baseline-conditioned residualization, which removed baseline-driven RTM components from both sides of every correlation, the diagonal had a mean of 0.52 (range 0.44–0.61) and the off-diagonal was 0.12). For raw change scores, within-domain prediction (matrix diagonal) increased to r = 0.56 (range 0.48–0.60), whereas cross-domain prediction (off-diagonal) was reduced to r = 0.06, confirming that retest-learning was domain-localized under either analytic framework. Memory and Processing Speed diagonal cells were essentially unchanged by baseline-conditioning (raw 0.56 and 0.59; baseline-residualized 0.56 and 0.61), consistent with their negligible RTM contributions documented in Table 2. Executive Function and Language/Story Memory diagonal cells showed greater attenuation (EF: 0.48 → 0.44; LS: 0.60 → 0.47), reflecting removal of the RTM and floor-recovery components from those domains.

Consistent with these findings, the five domains' overnight learning gains were themselves mutually uncorrelated (mean pairwise r = 0.03), and a principal-components analysis of the five-domain gain correlation matrix yielded near-flat eigenvalues with the first component explaining only 23% of variance — close to the 20% expected under full independence — indicating no general "good-retest-learner" factor. The prognostic value of the omnibus composite (r = 0.61) therefore arises from the aggregation of five largely independent domain-specific retest-learning signals rather than from a single underlying trait.

We also examined within-domain coherence at finer granularity using the 10 cluster scores derived from the CCAB bifactor decomposition (two clusters per domain). The two clusters within each domain showed only modest correlation in their overnight retest gains (r = 0.10–0.28 for LS, Memory, PS, SF; r ≈ 0.10 for EF; Cronbach's α = 0.18–0.44). In a 10×10 cluster-level cross-prediction matrix, each cluster's overnight gain predicted its own mean 6–30 month trajectory at r ≈ 0.56 (range 0.44–0.65) but predicted its within-domain partner cluster's trajectory at only r ≈ 0.14 on average, with cross-domain cluster prediction near chance (r ≈ 0.04). Retest learning was therefore largely localized to cluster-specific processes rather than reflecting unitary domain-level learning.

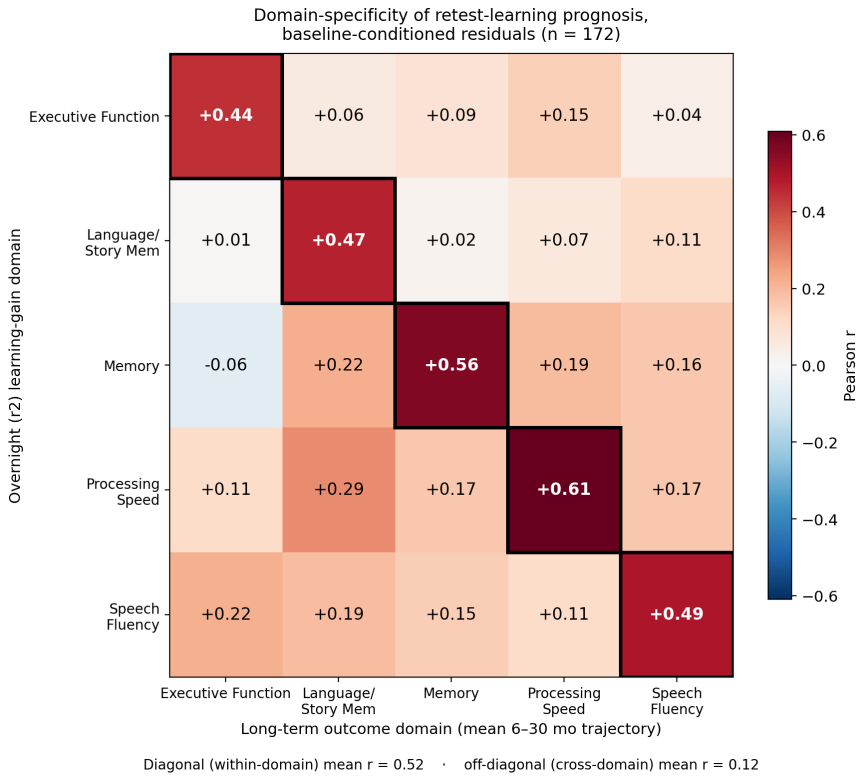

Figure 6. Domain specificity of retest learning effects under baseline-conditioned residualization (n = 172). Each cell is computed after residualizing both the row overnight gain and the column mean follow-up gain on the corresponding domain baseline. Within-domain prediction (boxed diagonal) averaged r = 0.52 (range 0.44–0.61); cross-domain prediction (off-diagonal) averaged r = 0.12, showing that retest-learning was primarily domain-localized rather than reflecting a general learning trait.

Clinical Implications

To examine whether the lowest tail of the retest-learning distribution carried clinical signal beyond quartile effects, we identified the 12 participants whose baseline and demographically adjusted omnibus overnight change fell in the bottom 7% of the cohort (residual threshold ≤ −0.35 z). These subjects gained substantially less than the cohort norm at overnight retest (mean adjusted retest-learning residual = −0.50 z, $t(11) = -12.4$, $p < 10^{-9}$), and showed persistent below-expectation residuals at every follow-up wave: 6 months = −0.30 z ($t(11) = -3.04$, $p = 0.011$), 18 months = −0.17 z ($t(11) = -2.72$, $p = 0.020$), 30 months = −0.24 z ($t(11) = -3.13$, $p = 0.010$). Bottom 7% participants did not differ from the rest of the cohort in age (71.0 vs 70.9 yr, $p = 0.97$), education ($p = 0.49$), vocabulary ($p = 0.35$), or omnibus baseline ($p = 0.94$), yet all 12 (100%) landed in the cohort's trajectory bottom quartile and 9 of 12 (75%) showed net adjusted decline.

Table 5 documents the prognostic value of overnight retest-learning using logistic regression to predict bottom-quartile cognitive trajectory (mean of 6, 18, and 30 month change) from overnight gain. This analysis was performed for omnibus scores and separately for each cognitive domain. Odds ratios are reported per 1-SD decrease in overnight gain so that larger OR values correspond to greater risk of Q1 membership. The omnibus composite predicted bottom-quartile 30-month trajectory at OR = 2.53 (95% CI 1.65–3.88) per 1-SD decrease in overnight gain (AUC = 0.72), rising to OR = 2.70 (95% CI 1.65–4.42, AUC = 0.82) after adjustment for baseline and demographic covariates. Per-domain ORs ranged from 2.27 (Executive Function, adjusted) to 4.21 (Memory, adjusted), with adjusted AUCs of 0.82–0.87. The omnibus composite produced an adjusted AUC of 0.82, below the strongest single domains (Memory and Speech Fluency, AUC 0.84–0.86), consistent with aggregation across the five domains improving precision beyond any single domain.

Domain	Gain SD (z)	Unadjusted OR per 1-SD ↓ [95% CI]	AUC	r1+demographics-adjusted OR per 1-SD ↓ [95% CI]	AUC
Omnibus Composite	**0.28**	**2.53 [1.65–3.88]**	**0.72**	**2.70 [1.65–4.42]**	**0.82**

Executive Function	0.54	2.31 [1.55–3.46]	0.72	2.27 [1.42–3.65]	0.83
Language/Story Memory	0.61	4.17 [2.40–7.26]	0.79	3.37 [1.70–6.67]	0.87
Memory	0.50	3.74 [2.25–6.21]	0.79	4.21 [2.31–7.68]	0.84
Processing Speed	0.45	2.72 [1.69–4.37]	0.73	2.79 [1.49–5.20]	0.87
Speech Fluency	0.62	4.35 [2.45–7.72]	0.79	4.16 [2.16–8.00]	0.86

Table 5. Continuous logistic odds ratios for predicting bottom-quartile mean 6–30 month cognitive trajectory from retest-learning gain (n = 172). For each domain, the predictor is retest gain in that domain and the outcome is the binary indicator of bottom-quartile mean long-term performance in the same domain. ORs are reported per 1-SD decrease in retest gain (the SD of gain differs by domain; see column 2), so that OR > 1 corresponds to elevated decline risk in the biomarker-conventional direction. Unadjusted models include only the gain term. Adjusted models corrected for baseline performance plus age, sex, vocabulary, education, race [Black, Asian, Other; White = reference], and research group [VA vs. NBS Normative]). AUC = area under the receiver operating characteristic curve. All ORs significant at p < 0.001 (unadjusted) and p < 0.01 (adjusted).

Discussion

These results extend the literature on short-term practice effects in four ways. First, we found that overnight retest-learning magnitude carried significant prognostic information for cognitive trajectories over 30 months — substantially longer than the 1- to 12-month intervals examined in most prior studies. Second, the prognostic signal was robust across analytic frameworks: it was evident in raw change scores, survived baseline-conditioned and demographic residualization, and persisted after explicit decomposition of regression-to-the-mean contributions. Third, we identified substantial domain heterogeneity in unresidualized prognostic magnitude (Q4–Q1 30-month gaps ranged from 0.29 z in Memory to 0.94 z in Language/Story Memory) that collapsed to a tighter band of 0.22–0.59 z after demographic and baseline adjustment, indicating that much of the apparent between-domain variation reflects demographically-mediated trajectory differences, including Language/Story floor-recovery dynamics, rather than intrinsic domain-level prognostic differences. Fourth, the prognostic signal was domain-specific: each domain's overnight gain predicted its own 30-month trajectory at r ≈ 0.56 but predicted other domains' trajectories at near-chance levels (r ≈ 0.05).

Domain-specific retest-learning

The differential magnitude of retest effects in different cognitive domains is broadly consistent with meta-analytic findings from the wider neuropsychological literature. Calamia and colleagues (2012) meta-analyzed practice effects across 1,597 individual effect sizes from studies of healthy adults and reported an overall average effect of approximately 0.25 SD, with substantial variation by domain: memory tests (particularly verbal list-learning and figural memory tasks) showed the largest practice effects, while measures of processing-speed and executive-control showed smaller effects. Scharfen et al. (2018), in a separate meta-analysis across 50 studies, similarly reported that memory tasks produced the largest gains across repeated administrations (d ≈ 0.5) while processing speed tasks showed smaller and more rapidly plateauing effects. These domain-specific patterns parallel our cohort-mean findings: Memory showed the largest overnight gain (+0.98 z) in the present cohort, while Executive Function showed the smallest (+0.15 z).

Domain differences reflect improvements in different cognitive operations. Memory practice effects are typically attributed to encoding and retrieval of specific test items — the participant remembers the actual words or stories from the prior administration. Processing speed practice effects reflect a different mechanism: gains primarily from procedural learning (motor sequencing, response-mapping familiarity) rather than item-specific recall. Executive function tasks, particularly those depending on inhibitory control, set-shifting, or working memory updating, show the smallest practice effects in the literature and are often considered insensitive to short-term retest gains because cognitive demands shift with realtime processing contingencies, even when overall test structure is preserved. Language and verbal fluency measures typically show intermediate retest effects (Calamia et al., 2012), often complicated by ceiling effects in high-functioning samples and floor-recovery dynamics in low-functioning samples, which align with the LS-domain patterns we observed.

Our findings extend this domain-specific literature in two ways. First, by demonstrating that overnight retest learning at 1 day (rather than 1 week) follows similar domain-magnitude rankings, suggesting that the

underlying mechanisms are engaged by overnight testing. Second, by tracking these domain-stratified signals over 30 months, we found that prognostic information carried by retest-learning is also domain-heterogeneous: Processing Speed and Language/Story Memory showed the largest sustained trajectory differentiation among learning-magnitude quartiles, while Memory — despite having the largest cohort-mean practice effect — showed more modest quartile differentiation. This dissociation between cohort-mean magnitude and between-subject prognostic value is methodologically important: a domain with universally strong learning effects is less useful for identifying individual differences than a domain where learning varies between subjects.

Retest learning in different domains

Memory shows the cleanest retest-learning signal in our cohort on psychometric grounds: high test-retest reliability (TRR = 0.88), large cohort-mean practice effects (+0.98 z) that persist at +0.59 z through 30 months, and minimal contribution from classical RTM (8%). Nonlinearity tests, however, revealed a significant inverted-U baseline-gain relationship in Memory, with subjects at high baseline gaining less than middle-baseline subjects. The hinge model localized this to a ceiling-suppression term ($\beta = -0.64$, $p < 0.001$) with no corresponding floor-recovery term — consistent with a measurement ceiling constraining gain magnitude in the most capable subjects rather than an RTM artifact. Because ceiling effects compress rather than inflate the observed gain, the prognostic correlation reported here is a conservative estimate of what would be observed in the absence of ceiling-bound subjects. This makes Memory the most defensible single-domain retest-learning indicator on psychometric grounds, consistent with the dominance of memory measures in Duff's protocols (Duff et al., 2014, 2017).

Language/Story Memory showed large initial practice effects (+0.77 z) and the largest sustained Q4-vs-Q1 trajectories gap at 30 months (0.94 z in raw change scores), but its baseline correlated with gain more than classical RTM predicts (% RTM = 131%). We interpret this as evidence of floor-recovery dynamics: subjects who performed poorly at baseline on novel story-memory material — likely due to encoding and retrieval failures on first exposure, with hesitant speech and poor lexical structure — showed disproportionately large gains at retest when re-exposed to the same material. This is mechanistically distinct from classical RTM, which assumes a stable true score with random measurement error. The clinical implication is that LS retest-learning may index task-familiarity acquisition as much as underlying cognitive capacity, and should be interpreted with appropriate caution.

Perhaps the most theoretically informative result was that retest-learning was domain-specific rather than a general trait (Figure 6). Each domain's overnight gain predicted that same domain's 30-month trajectory at $r \approx 0.56$, with chance prediction of the trajectories of other domains (mean cross-domain $r = 0.05$). This pattern argues against a nonspecific practice mechanism — such as a global increase in task engagement — which would produce correlated gains across domains. Instead, the localization of each prognostic signal to its own cognitive domain implies that overnight retest-learning gains reflect the integrity of separable underlying neural systems: episodic-memory consolidation circuitry for the Memory and Language/Story domains, fronto-striatal and white-matter efficiency for Processing Speed, and language-network integrity for Speech Fluency and Lexical/Story processing. Thus, retest-learning is best understood not as a single "learning ability" but as a family of system-specific capacities, each reflecting the neural circuitry that supports that domain.

This framing connects the present findings to the cognitive-reserve and brain-maintenance literatures. Cognitive reserve refers to the capacity to maintain function in the face of accumulating age- or disease-related brain change, and is classically indexed by proxies such as education, occupational complexity, and vocabulary (Stern, 2012). Overnight retest-learning provides a different, performance-based index of the same latent construct: rather than inferring reserve from lifetime exposures, it measures the brain's actual capacity to encode, consolidate, and re-express new learning. The observation that vocabulary — a cognitive reserve proxy — independently predicted more favorable long-term trajectories in our cohort (Table 4), alongside and separately from retest-learning, is consistent with retest-learning and traditional reserve proxies capturing related but non-redundant aspects of adaptive capacity. Domain-specific retest-learning may thus operationalize a domain-resolved form of reserve, capturing latent system integrity, residual plasticity, or subclinical disease burden at the level of individual cognitive systems rather than as a single global quantity.

The domain-specific structure also carries a methodological lesson: global composite scores can obscure separable trajectories across cognitive systems. The omnibus composite is a powerful prognostic instrument precisely because it aggregates five independent domain-specific signals, reducing domain-

specific noise; but it discards information about which domain is changing. Two individuals with identical omnibus retest-learning scores may have entirely different underlying profiles — one consolidating episodic memory with little processing-speed plasticity, the other the reverse — with potentially different prognostic signals for the different clinical syndromes (amnestic versus dysexecutive or speed-predominant). The dissociation we observed between cohort-mean practice-effect magnitude and between-subject prognostic value reinforces this point: a domain in which everyone improves similarly (high mean gain, low variance) carries little individual-differences information, whereas a domain with substantial between-subject variation in learning is prognostically informative even when its average gain is modest. Reporting domain-resolved retest-learning alongside the composite therefore preserves clinically meaningful heterogeneity that a single global score would mask.

Magnitude and persistence of the prognostic signal

The cohort-wide correlation of r = 0.61 between overnight retest learning and 30-month trajectories exceeds typical prognostic correlations reported for established AD biomarkers — plasma p-tau217 (r ≈ 0.3–0.4), CSF Aβ42/Aβ40 (r ≈ 0.2–0.4), or hippocampal volume (r ≈ 0.2–0.3) — for cognitive trajectories in cognitively unimpaired older adults. The fact that Q1 subjects fell below baseline at 6 months and remained there for the next two years — while Q4 subjects retained approximately half of their initial overnight gain indefinitely — represents a significant dose-response pattern that supports the interpretation of 1-day retest-learning as a stable trait-like indicator of cognitive integrity.

Candidate neural mechanisms

The cognitive integrity indexed by overnight retest-learning likely reflects multiple neural mechanisms rather than a single process. Hippocampal integrity is the most prominent single candidate. The overnight interval spans a complete sleep cycle and the systems-consolidation window during which newly encoded declarative memories are stabilized through hippocampal-cortical replay during slow-wave sleep (Klinzing et al., 2019). Subtle reductions in hippocampal volume — below the threshold required to produce a measurable cross-sectional memory deficit — have been shown to attenuate short-term learning in cognitively unimpaired older adults, with the largest effects in amyloid-positive subjects (Lim et al., 2020). This mechanism fits the present domain pattern, in which Memory shows the largest 1-day cohort-mean gain and the most persistent long-term elevation. Sleep architecture, particularly slow-wave sleep, provides the temporal window during which hippocampal-cortical replay occurs, and between-subject variation in slow-wave sleep duration and intensity is substantial in older adults (Mander et al., 2017). Sleep apnea, sleep fragmentation, and primary insomnibusia each attenuate overnight consolidation and could plausibly account for a meaningful fraction of the variance in our retest-learning signal. Default mode network (DMN) connectivity has been linked to memory consolidation and is among the earliest functional changes downstream of amyloid deposition in cognitively unimpaired older adults (Sperling et al., 2009), making DMN integrity a candidate substrate for the prognostic signal observed here.

Additional candidates include basal forebrain cholinergic tone, which supports attention-mediated encoding and is degraded early in AD pathology, sometimes before hippocampal atrophy is detectable on MRI (Grothe et al., 2012) and chronic low-grade neuroinflammation, which impairs hippocampal long-term potentiation through microglial-mediated synaptic dysfunction. Genetic factors including APOE ε4 (Egan et al., 2003) — likely explain additional variance in consolidation efficiency

Implications for clinical trials and prognostic screening

The findings suggest several potential applications in trial design and longitudinal cognitive monitoring. The strong association between overnight retest-learning and 30-month trajectories raises the possibility that retest-learning could serve as a stratification variable for trial enrichment. Trials seeking to enroll likely decliners might preferentially enrich for bottom-quartile retest learners, who in our cohort returned to baseline by 6 months and remained there through 30 months. The overnight retest is operationally simple compared to fluid or imaging biomarkers and can be administered at home, which may make it attractive as a low-cost first-stage screen prior to more invasive or expensive biomarker assessment. In OR-and-AUC terms, a 1-SD decrease in omnibus overnight gain conferred OR = 2.70 (95% CI 1.65–4.42, AUC = 0.82) after adjustment for baseline and ten demographic covariates (Table 5). For comparison, plasma p-tau217 alone yields AUCs in the 0.78–0.82 range for cognitive decline prediction, hippocampal MRI volume 0.75–0.80, and combined plasma-plus-amyloid-PET panels 0.88–0.92; the AUC = 0.82 we report here is

comparable to single-modality blood-based or imaging biomarkers and approaches the predictive value of combined multi-modal panels. In addition, retest-learning may alter treatment-response profiles. The mechanisms most plausibly underlying retest-learning heterogeneity — hippocampal integrity, sleep-dependent consolidation, cholinergic tone, neuroinflammation — are precisely the mechanisms targeted, directly or indirectly, by several classes of investigational therapies in cognitive aging and early AD.

Because all participants were classified as cognitively unimpaired at enrollment and formal MCI/AD adjudication was not collected, this subgroup is best characterized as exhibiting a candidate prodromal cognitive signature: minimal-or-negative learning combined with persistent below-expectation trajectories across 30 months, demographically indistinguishable from cognitively typical peers but identifiable through a single 1-day retest.

Limitations

Sample selection. The cohort was recruited in the San Francisco Bay Area and was substantially more educated (mean ~16.6 years) than the general U.S. population of older adults, with the NBS community subsample in particular averaging ~17.4 years (some graduate education). This reflects the well-documented self-selection of higher-reserve, more health-literate individuals into longitudinal cognitive-aging research. Because higher educational attainment and vocabulary predicted more favorable long-term trajectories in our data, a high-reserve sample may render the present demonstration of retest-learning prognostic value conservative relative to what would be observed in a more representative, lower-reserve population. In addition, the two contributing subsamples — military veterans (VA) and community volunteers (NBS) — differed substantially in sex, age, education, and vocabulary (Table 1), and neither is demographically representative of the broader older-adult population. Although we covaried research group in all demographic-adjustment analyses, replication in population-based and demographically representative cohorts is needed before the overnight retest-learning metric can be applied as a prognostic tool at scale.

Adherence. The analysis was restricted to the 172 participants who completed all five assessment waves through the 30-month follow-up. This restriction reflected calendar-time enrollment — these were the earliest enrollees in the ongoing CCAB normative study, who had reached the 30-month visit — rather than selection on adherence or cognitive status. Excluded participants were predominantly partial completers still progressing through the follow-up schedule rather than dropouts. Retest-learning trajectories computed in the larger 6 month ($N \approx 258$) and 18 month ($N \approx 200$) subsamples were similar to those reported here, arguing against substantial completer bias. Nonetheless, participants who adhere to a demanding multi-year testing schedule may differ systematically from the general population of older adults — for example, in health status, motivation, conscientiousness, or cognitive reserve — and such factors could plausibly be correlated with both retest-learning capacity and long-term cognitive trajectory. To the extent that adherent completers are healthier and higher in reserve than the broader population, the present cohort may underrepresent the rapid-decliner end of the trajectory distribution, which would attenuate rather than inflate the observed prognostic separation.

Conclusions

A single 1-day retest of the CCAB battery generated retest-learning measures that predict cognitive trajectories over 30 months in cognitively unimpaired older adults. Aggregate retest learning reflected domain-specific processes, was robust to multiple analytic frameworks, and showed a cohort-wide correlation ($r = 0.61$) with cognitive trajectory that exceeds the prognostic correlations reported for established AD biomarkers in cognitively unimpaired older adults. Memory and Processing Speed carried the cleanest retest-learning signals. These findings support the use of 1-day CCAB retest as a low-burden, scalable prognostic tool for trial enrichment and longitudinal cognitive monitoring.

Declarations

Funding. This work was supported by NIA R44AG097322 (PI: David L. Woods). The funder had no role in study design, data collection, analysis, interpretation, or manuscript preparation.

Competing interests. All authors are employees or collaborators with Neurobehavioral Systems, Inc., the developer of the CCAB battery.

Ethics approval. Study procedures were approved by WIRB protocol 20201196. All participants provided written informed consent.

Clinical trial registration. ClinicalTrials.gov NCT04800588.

Data availability. De-identified data and analysis code are available for reasonable academic requests to the corresponding author.

Author contributions. All authors contributed to study design, data analysis, manuscript drafting, and approval of the final version.

References

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. The Clinical Neuropsychologist, 26(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. Neuropsychology, 7(1), 41–52. https://doi.org/10.1037/0894-4105.7.1.41

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. Cognitive Neuropsychology, 24(4), 343–372. https://doi.org/10.1080/02643290701290146

Duff, K., Foster, N. L., & Hoffman, J. M. (2014). Practice effects and amyloid deposition: Preliminary data on a method for enriching samples in clinical trials. Alzheimer Disease & Associated Disorders, 28(3), 247–252. https://doi.org/10.1097/WAD.0000000000000021

Duff, K., Hammers, D. B., Dalley, B. C. A., Suhrie, K. R., Atkinson, T. J., Rasmussen, K. M., … Hoffman, J. M. (2017). Short-term practice effects and amyloid deposition: Providing information above and beyond baseline cognition. The Journal of Prevention of Alzheimer's Disease, 4(2), 87–92. https://doi.org/10.14283/jpad.2017.9

Duff, K., Suhrie, K. R., Dalley, B. C. A., Anderson, J. S., & Hoffman, J. M. (2019). External validation of change formulae in neuropsychology with neuroimaging biomarkers: A methodological recommendation and preliminary clinical data. The Clinical Neuropsychologist, 33(3), 478–489. https://doi.org/10.1080/13854046.2018.1484518

Duff, K., Miller, J. B., Cobos, K., Rodrigues, J., & O'Bryant, S. E. (2024). Derivation of indices of cognitive change among Hispanic adults and elders. JAMA Network Open, 7(9), e2431180. https://doi.org/10.1001/jamanetworkopen.2024.31180

Egan, M. F., Kojima, M., Callicott, J. H., Goldberg, T. E., Kolachana, B. S., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M., Lu, B., & Weinberger, D. R. (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. Cell, 112(2), 257–269. https://doi.org/10.1016/S0092-8674(03)00035-7

Farrell, M. E., Kennedy, K. M., Rodrigue, K. M., Wig, G., Bischof, G. N., Rieck, J. R., Chen, X., Festini, S. B., Devous, M. D., Sr., & Park, D. C. (2017). Association of longitudinal cognitive decline with amyloid burden in middle-aged and older adults: Evidence for a dose-response relationship. JAMA Neurology, 74(7), 830–838. https://doi.org/10.1001/jamaneurol.2017.0892

Grothe, M., Heinsen, H., & Teipel, S. J. (2012). Atrophy of the cholinergic basal forebrain over the adult age range and in early stages of Alzheimer's disease. Biological Psychiatry, 71(9), 805–813. https://doi.org/10.1016/j.biopsych.2011.06.019

Hammers, D. B., Suhrie, K. R., Dixon, A., Porter, S., & Duff, K. (2021). Reliable change in cognition over 1 week in community-dwelling older adults: A validation and extension study. Archives of Clinical Neuropsychology, 36(3), 347–358. https://doi.org/10.1093/arclin/acz076

Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., & Morris, J. C. (2015). Absence of practice effects in preclinical Alzheimer's disease. Neuropsychology, 29(6), 940–948. https://doi.org/10.1037/neu0000208

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. Nature Neuroscience, 22(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Lim, Y. Y., Baker, J. E., Bruns, L., Jr., Mills, A., Fowler, C., Fripp, J., Rainey-Smith, S. R., Ames, D., Masters, C. L., & Maruff, P. (2020). Association of deficits in short-term learning and Aβ and hippocampal volume in cognitively normal adults. Neurology, 95(18), e2577–e2585. https://doi.org/10.1212/WNL.0000000000010728

Mander, B. A., Winer, J. R., & Walker, M. P. (2017). Sleep and human aging. Neuron, 94(1), 19–36. https://doi.org/10.1016/j.neuron.2017.02.004

Ossenkoppele, R., Salvadó, G., Janelidze, S., Pichet Binette, A., Bali, D., Karlsson, L., Palmqvist, S., Mattsson-Carlgren, N., Stomrud, E., Therriault, J., Rahmouni, N., Rosa-Neto, P., Coomans, E. M., van de Giessen, E., van der Flier, W. M., Teunissen, C. E., Jonaitis, E. M., Johnson, S. C., Villeneuve, S., … Hansson, O. (2025). Plasma p-tau217 and tau-PET predict future cognitive decline among cognitively unimpaired individuals: Implications for clinical trials. Nature Aging, 5(5), 883–896. https://doi.org/10.1038/s43587-025-00835-z

Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. Intelligence, 67, 44–66. https://doi.org/10.1016/j.intell.2018.01.003

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. Assessment, 25(1), 14–30. https://doi.org/10.1177/1073191116643164

Sperling, R. A., Laviolette, P. S., O'Keefe, K., O'Brien, J., Rentz, D. M., Pihlajamaki, M., Marshall, G., Hyman, B. T., Selkoe, D. J., Hedden, T., Buckner, R. L., Becker, J. A., & Johnson, K. A. (2009). Amyloid deposition is associated with impaired default network function in older persons without dementia. Neuron, 63(2), 178–188. https://doi.org/10.1016/j.neuron.2009.07.003

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. The Lancet Neurology, 11(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Tort-Merino, A., Pérez-Millan, A., Falgàs, N., Borrego-Écija, S., Guillén, N., Sarto, J., Esteller, D., Bosch, B., Castellví, M., Juncà-Parella, J., Val-Guardiola, A., Fernandez-Villullas, G., Antonell, A., Sánchez-Saudinós, M. B., Rubio-Guerra, S., Zhu, N., García-Martínez, M., Pozueta, A., Estanga, A., … Sánchez-Valle, R. (2024). Decreased practice effects in cognitively unimpaired amyloid-beta positive individuals: A multicenter, longitudinal, cohort study. Alzheimer's & Dementia, 21(3), e70016. https://doi.org/10.1002/alz.70016

Woods, D. L., Pebler, P., Johnson, D. K., Herron, T., Hall, K., Blank, M., Geraci, K., Williams, G., Chok, J., Lwi, S., Curran, B., Schendel, K., Spinelli, M., & Baldo, J. (2024). The California Cognitive Assessment Battery (CCAB). Frontiers in Human Neuroscience, 17, 1305529. https://doi.org/10.3389/fnhum.2023.1305529

Woods, D. L., Hall, K., Jaramillo, I., Blank, M., Geraci, K., Pebler, P., Cole, M., & Johnson, D. K. (2026a). Comprehensive demographic correction in cognitive assessment: A LASSO-based normative scoring framework for the California Cognitive Assessment Battery. In preparation.

Woods, D. L., Hall, K., Jaramillo, I., Blank, M., Geraci, K., Pebler, P., Cole, M., & Johnson, D. K. (2026b). Bifactor confirmatory factor analysis of the California Cognitive Assessment Battery reveals separable cognitive domains and a general factor. In preparation.

Zheng, B., Udeh-Momoh, C., Watermeyer, T., de Jager Loots, C. A., Ford, J. K., Robb, C. E., Giannakopoulou, P., Ahmadi-Abhari, S., Baker, S., Novak, G. P., Price, G., & Middleton, L. T. (2022). Practice effect of repeated cognitive tests among older adults: Associations with brain amyloid pathology and other influencing factors. Frontiers in Aging Neuroscience, 14, 909614. https://doi.org/10.3389/fnagi.2022.909614