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 Alzheimer's Disease Research Center, Walnut Creek, CA

Introduction

Performance usually improves when cognitive tests are retaken, and the magnitude of improvement may predict Alzheimer's disease pathology in older adults (Duff et al., 2014). We tested 172 healthy older adults (mean age 72 years) on consecutive days to measure overnight retest learning, then retested them at 6, 18, and 30 months to measure cognitive trajectory. Subjects with >1 SD below-average overnight learning gain had 4x the odds of being in the lowest 30-month cognitive trajectory group.

Methods

Overview. We examined the cognitive trajectories of 172 normal older adults (mean age 71 years at enrollment) on 15 subtests of California Cognitive Assessment Battery (CCAB) administered at intervals of one day, 6 months, 18 months, and 30 months. The study included two cohorts: 90 military veterans recruited through the Veterans Administration (VA) and 82 community-dwelling volunteers recruited by Neurobehavioral Systems. All participants were tested telemedically in their homes or in a laboratory setting (Woods et al., 2024).

Tests, scores, and cognitive domains. The 15 CCAB subtests generated 62 measures including Speech and Language Biomarkers (SLBs). Scores were analyzed individually and grouped into five cognitive domains: Executive Function (EF), Processing Speed (PS), Memory (EM), Language/Story Memory (LS), and Speech Fluency (SF). Each domain contained 7-16 measures.

Quartile Stratification. Subjects were stratified into quartiles (Q1-Q4) ordered by the magnitude of overnight retest learning gain in the omnibus score.

Trajectory Metrics. Two complementary trajectory metrics were computed at each wave. Raw change scores were defined per subject as (wave z − enrollment z), without residualization. Demographically corrected baseline-conditioned residuals were obtained from cohort-level OLS regressions predicting each follow-up z-score from enrollment performance (Chelune et al., 1993) with corrections for age, education, sex, race, vocabulary, and recruitment cohort.

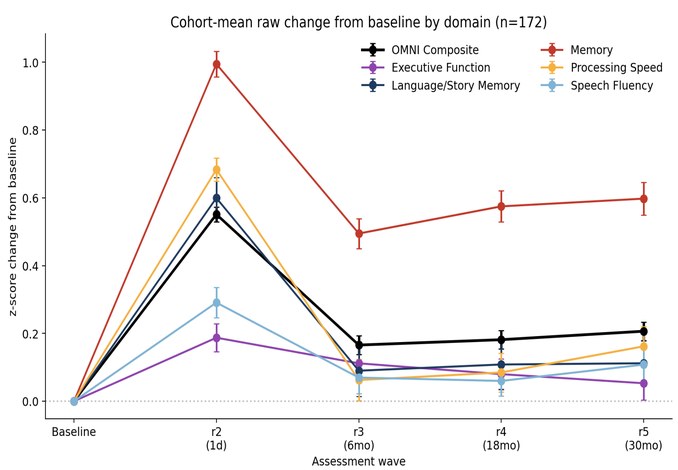

Figure 1. Overnight retest learning effects in different cognitive domains. Participants showed substantial overnight retest learning effects on the omnibus composite (black line, 0.55 z, t(171) = 25.1, p < 10^{-62}) that decayed substantially by 6 months and stabilized for the remaining 24 months. Learning differed substantially by cognitive domain: Memory (red line) showed the largest 1-day practice effect (+0.99 z, t(171) = 26.3) and the most preserved long-term elevation (30 months = +0.60 z, t(171) = 12.3). Language/Story Memory and Processing Speed also showed substantial overnight learning but more modest retention, while Speech Fluency and Executive Function showed minimal learning that had largely returned to baseline by 6 months.

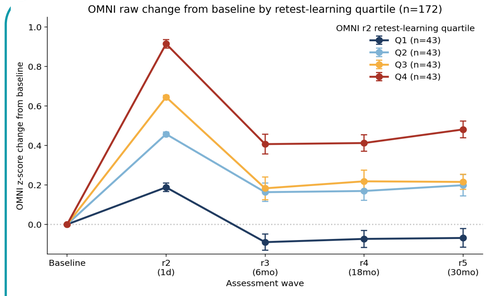

Figure 2. Learning quartiles predict long-term trajectories. Unresidualized omnibus z-scores did not differ in baseline performance (ANOVA : F(3,168) = 0.72, p = 0.54) but diverged sharply in the magnitude of overnight learning. Shown relative to baseline, (Q1) gained +0.19 z, Q2 +0.46 z, Q3 +0.64 z, and Q4 +0.91 z. Between six and 30 months Q2-Q4 regressed toward the mean, with rank-ordering preserved at subsequent waves. However, Q1 dropped slightly below baseline -0.07 and remained there.

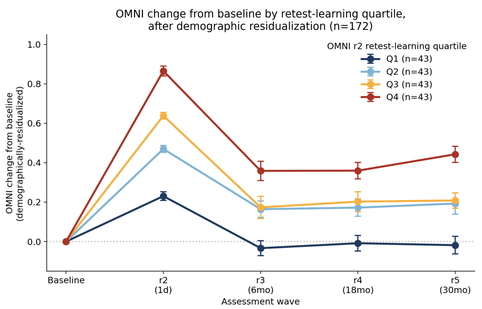

Figure 3. Residualized quartiles corrected for demographic influences. Omnibus z-scores were calculated relative to performance at enrollment after correction for demographic influences of age, education, gender, race, vocabulary, and research group. No demographic factors showed a significant influence on retest learning magnitude, but higher vocabulary scores were marginally associated with improved 30-month outcome (p = 0.04).

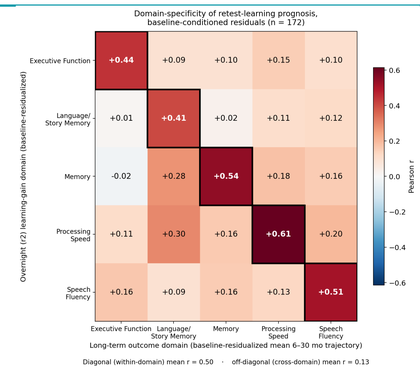

Figure 4. Domain specificity of retest learning. Domain-specificity of retest-learning prediction for residualized scores (n = 172). Each cell shows the Pearson correlation between overnight retest-learning gain (row) and the mean 6–30 month trajectory (column). Same-domain prediction (boxed diagonal) averaged r = 0.50 (range 0.41–0.61); cross-domain prediction averaged r = 0.13.

Summary

- Participants whose overnight learning was > 1 SD below the cohort mean showed four-fold greater odds of belonging to the worst long-term trajectory group (adjusted OR = 4.26, AUC = 0.87).
- Learning was domain-specific: improvement in memory predicted future memory, etc., with gains in one cognitive domain independent from gains in others.
- Retest-learning magnitude correlated strongly with mean 6–30 month cognitive trajectory (r = 0.60).
- Learning-stratified quartiles had similar baseline performance but markedly different long-term outcomes. Participants with the largest overnight gains retained much of that benefit across follow-up, whereas those with the smallest gains returned to baseline by 6 months.
- The correlation between overnight learning magnitude and cognitive outcomes was similar to that of combined p-tau217 and PET imaging.
- Reduced overnight learning identifies higher-risk individuals to flag patients who warrant closer follow-up and enrich clinical trials.

Discussion

The cohort-wide correlation of r = 0.60 between overnight retest learning and 6 to 30-month cognitive performance in cognitively unimpaired old adults is similar to the predictive accuracy of many 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). The fact that Q1 subjects fell below baseline by 6 months and remained there for the next two years — while Q4 subjects retain approximately half of their initial 1-day gain for 2.5 years — represents a significant dose-response pattern that supports the interpretation of 1-day retest-learning as a stable marker of cognitive resilience.

The magnitude of overnight learning in the OMNI composite measure predicted membership in the lowest trajectory quartile at 30 months at OR = 3.91 (95% CI 2.33–6.52) per 1-SD decrease in overnight gain (AUC = 0.80), rising to OR = 64.26 (95% CI 2.35–7.73, AUC = 0.87) after adjustment for baseline and demographic covariates. Demographically corrected per-domain ORs ranged from 2.27 (Executive Function) to 3.65 (Language/Story Memory), with adjusted AUCs of 0.81–0.88. The OMNI composite showed the highest adjusted AUC (0.87), approaching the predictive accuracy of combined p-tau217 and PET neuroimaging (Ossenkoppele et al. 2025).

References

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.

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).*

Ossenkoppele, R., Salvadó, G., Janelidze, S., Pichet Binette, A., Bali, D., Karlsson, L., Palmqvist, S., Mattsson-Carlgren, N., Stomrud, E., Therriault, J., Rahmouni, et. al. (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.

Woods, D., 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, 18*, 1305529.

Contact us

drdlwoods@neurobs.com
ccabresearch.com
neurobs.com

Visit us at booth 1417

Supported by NIA R44AG062076, NIA R44AG080951