Factor Analysis of Comprehensive Cognitive Testing Reveals Age-related Cognitive Dedifferentiation

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

Supported by NIA R44AG097322.

Correspondence: drdlwoods@neurobs.com.

Short title: CCAB Factor Analysis and Cognitive Dedifferentiation

Keywords: computerized neuropsychological assessment; bifactor model; cognitive dedifferentiation; Spearman's Law of Diminishing Returns; cognitive aging; individual differences; demographically adjusted scores

Key Findings

- We evaluated age-related dedifferentiation of latent cognitive domains in 1916 volunteer adults aged 18 to 89, comparing the general factor (g) with domain composites in five cognitive domains: Executive Function (EF), Lexical/Story processing (LS), Memory (EM), Processing Speed (PS), and Speech Fluency (SF).
- We examined dedifferentiation with two bifactor models using (1) unregressed indicators (correlated with age in many cases), and (2) residualized indicators corrected for age and other demographic predictors. Consistent with cognitive dedifferentiation g-loadings on cognitive domain scores increased with age in both datasets.
- Residualized domain composites and bifactor factor scores correlated at r = 0.92–0.97 across age strata and ranked participants similarly on each dimension. However, the two scoring approaches diverged sharply in the lower tails of the distributions: factor scores flagged more domain-specific decrements than domain composite scores.
- In the age ≥ 65 subsample (N = 620), 23.1% fell in the bottom 7% of z-scores on g or at least one specific factor, distributed across different profile groups: Isolated g (2.7%), Isolated EM (3.4%), Isolated LS (3.9%), Isolated PS (3.1%), Isolated SF (2.9%), g+ (g with ≥1 factor; 4.4%), and g−, multi-factor (3.7%).
- Across the full sample, 78–84% of specific-factor deficits occurred in participants with intact general cognitive ability—a structural feature of bifactor parameterization that enhances early detection of focal cognitive change in high-functioning older adults that was less visible in domain-composite analysis.
- Among participants with g in the lowest quintile, domain composites overestimated domain-specific latent factor abnormalities by nearly 50%, while among participants with g in highest quintile domain composites underestimated domain-specific latent factor abnormalities by more than 80%.

Abstract

Background. Bifactor confirmatory factor analysis partitions cognitive test performance into a general factor (g) and orthogonal domain-specific factors. The clinical value of this decomposition reflects the magnitude of the g influence across cognitive domains which increases with age in a phenomenon known as cognitive dedifferentiation. Dedifferentiation is generally evaluated using unresidualized test scores from participants of different ages. However, uncorrected scores include the direct influence of age on indicators (e.g., motor slowing, reductions in peripheral nerve condition velocity) that introduce cross-domain loadings complicating the interpretation of dedifferentiation. Here we analyzed dedifferentiation in bifactor models using both uncorrected and demographically residualized scores in 1,916 normal subjects tested with the California Cognitive Assessment Battery (CCAB) and evaluated the impact of dedifferentiation on score interpretation in older subjects.

Methods. Bifactor confirmatory factor scores derived from 96 indicators in 1,916 native English-speaking adults (Woods et al., 2026, companion manuscript) were compared with five domain composites: Executive Function (EF), Lexical/Story processing (LS), Memory (EM), Processing Speed (PS), and Spoken Fluency (SF). For each domain, we quantified the proportion of composite variance attributable to g across six age strata (full sample, 18–59, ≥60, ≥65, ≥70, ≥75). Profile group composition in the age ≥ 65 subsample (N = 620) was characterized using an exploratory Mild Cognitive Impairment (MCI, bottom-7%) threshold applied independently to each of the five residualized factor scores, with 2,000-iteration bootstrap resampling, and threshold sensitivity analyses at 5% and 10%.

Results. Bifactor g-loadings on each cluster parcel rose substantially with age across age bins. A converging variance-decomposition analysis showed that the r^2 between each demographically residualized domain composite and g rose monotonically across cumulative age strata. Across all strate, domain composites and bifactor factor scores correlated at r = 0.92–0.97, with the two scoring approaches ranking participants similarly on each dimension. However, for low performing subjects bifactor scores showed increased domain selectivity compared to domain composites. In the age ≥ 65 subsample (N = 620), 23.1% fell below the bottom-7% threshold on g or ≥1 specific factor, partitioned into mutually exclusive groups: Isolated g (4.2%), Isolated EM (3.4%), Isolated LS (3.2%), Isolated PS (2.9%), Isolated SF (1.8%), g+ (3.9%), and g−, multi-factor (3.7%). Across the full sample, 78–84% of selective specific-factor deficits occurred in participants with intact general cognitive ability, a structural feature of bifactor parameterization not visible in domain composites. Among participants with g in the lowest quintile, domain composites overestimated domain-specific latent factor abnormalities by nearly 50%, while among participants with g in highest quintile domain composites underestimated domain-specific latent factor abnormalities by more than 80%.

Conclusions. Cognitive dedifferentiation increased the g share of domain-composite variance from young adults to age ≥75, roughly doubling its contribution to memory and lexical/story processing domains. Domain composites and bifactor factor scores ranked participants similarly across the lifespan, but the bifactor decomposition isolated domain-specific decrements in low performing subjects by removing the g component that increasingly dominated domain composite scores in older subjects.

Introduction

Factor analysis has shaped cognitive assessment for nearly a century. From Spearman's general factor through Cattell-Horn-Carroll (CHC) theory and the bifactor models that dominate contemporary psychometric work, the field has converged on the view that observed test performance reflects a hierarchical structure in which a general factor (g) coexists with separable specific abilities (Schneider & McGrew, 2018; Reise, 2012). The clinical translation of this structure into individual-patient interpretation is well established in the Wechsler scales, WAIS-IV (Wechsler, 2008) and WAIS-5, whose publisher-derived four- and five-factor index structures come directly from confirmatory factor analyses of standardization samples (Weiss, Keith, Zhu, & Chen, 2013), and in the Woodcock-Johnson IV (Schrank, McGrew, & Mather, 2014) and Kaufman Assessment Battery for Children, Second Edition (KABC-II; Reynolds, Keith, Fine, Fisher, & Low, 2007). Factor-derived composites are also primary cognitive endpoints in major aging research programs: the NIH Toolbox Cognition Battery reports Crystallized, Fluid, and Total cognition composites (Heaton et al., 2014); the Alzheimer's Disease Neuroimaging Initiative composite scores ADNI-Mem and ADNI-EF are standard cognitive outcomes in AD natural-history work (Crane et al., 2012; Gibbons et al., 2012); the Preclinical Alzheimer's Cognitive Composite (PACC) is the FDA-recognized cognitive endpoint in AD secondary-prevention trials (Donohue et al., 2014); and the NACC Uniform Data Set Version 3 (UDS-3) supports a published higher-order factor structure with demographic-adjustment calculator (Kiselica, Webber, & Benge, 2020), a derived executive composite (Staffaroni et al., 2021), and a more recent 4-factor model with measurement invariance across ethno-racial and linguistic groups (Matusz et al., 2025).

Cognitive dedifferentiation, the proportion of variance in test scores attributable to g, increases with age. Meta-analytic evidence across 408 studies spanning roughly a century of work supports both the age-related differentiation hypotheses (Blum & Holling, 2017) with effect sizes of practical significance (Breit, Brunner, Molenaar, & Preckel, 2022). Direct longitudinal evidence of dedifferentiation comes from the Berlin Aging Study, in which Ghisletta and Lindenberger (2003) documented age-based dedifferentiation of perceptual speed and crystallized knowledge in old age, and Tucker-Drob (2009) applied nonlinear factor analytic models to a nationally representative U.S. sample of 6,273 individuals and found graded ability-related differentiation across the lifespan. At the neural level, the converging evidence is equally strong: ventral visual cortex shows reduced functional specialization for object categories in older adults (Park et al., 2004), and the broader neural-dedifferentiation hypothesis — that age-related changes reduces the distinctiveness of neural networks and contribute to cognitive decline — is supported by multivariate pattern analysis across sensory, motor, and prefrontal regions (Koen & Rugg, 2019).

Cognitive dedifferentiation has a direct and important implication for clinical interpretation. In older subjects, g accounts for an increasing share of variance in domain test scores so that simple domain composites — the mean of several tests assigned to a putative cognitive domain — increasingly conflate general cognitive ability with domain-specific performance. Two consequences follow. First, domain composites become progressively less informative about domain-specific cognitive operations as participants age. Second, because the g component increasingly obscures the domain-specific deficit, selective domain deficits in high-functioning older adults become progressively harder to detect.

Bifactor parameterization partitions shared variance into a general factor and specific factors orthogonal to g, to isolate index domain-specific latent factors after the g contribution has been removed. The magnitude of the g-correction is therefore proportional to the degree of dedifferentiation. In samples where g and specific factors are well-differentiated (young adults), bifactor and domain composites segregate cleanly. However, in older samples where dedifferentiation has occurred, bifactor decomposition will potentially improve the identification of specific latent factor deficits by removing the increased contribution of g that contaminates domain composite scores.

Routine application of bifactor decomposition to individual-patient interpretation in adult neuropsychology has been limited by several structural and technological constraints of manual test administration. First, stable factor-score estimates require multiple indicators per factor. Typically. manual batteries obtain only 1-2 indicators per domain, leaving each putative latent factor with a narrow empirical base. Second, demographic confounding at the indicator level is typically addressed downstream of the factor solution rather than during indicator construction. For example, Kiselica et al. (2020) extracted factor scores on unregressed UDS-3 indicators and then applied demographic adjustment to the resulting factor scores. This approach does not

address the extent to which the recovered factor structure reflects shared demographic variance (e.g., age effects) loading heterogeneously across observed indicator scores. Third, manual neuropsychological administration introduces systematic between-site variation that is difficult to quantify or correct. Different examiners deliver test items at different rates, with different speech intensities and prosodic patterns, and apply scoring rules with variable strictness. These examiner-to-examiner differences appear as additional variance in the normative data and as noise in individual scores. When norms developed at one site are applied to performance acquired at another site, the examiner-source variance becomes an uncontrolled confound on individual-patient interpretation.

Computerized administration overcomes these constraints. The California Cognitive Assessment Battery (CCAB; Woods et al., 2024) administers 24 cognitive subtests in approximately 2.5 hours via telemedical proctoring, generating 96 process-level performance scores including accuracy, response time, kinematic features of motor output, acoustic-temporal speech markers (speaking rate, articulation rate, pause ratio, inter-word interval), and lexical-statistical features of speech (Honoré's statistic, word entropy, content keyword retrieval). With multiple indicators per cognitive domain, the CCAB supports stable factor-score estimation that manual administration cannot.

Here, we examine dedifferentiation in a large normative sample (n = 1916) using two bifactor models (Woods et al., 2026, companion manuscript) fit to two datasets: (1) unresidualized indicators clustered into cognitive domains and (2) indicators residualized with a Comprehensive scoring model that included age, age^2 residual, education, vocabulary, sex, race/ethnicity, computer use, and daily medications to corrrect demographic confounding at the indicator level. Similar bifactor architectures were seen in both datasets comprising a general factor (g) absorbing Executive Function (EF) variance plus four orthogonal specific factors — Processing Speed (PS), Memory (EM), Lexical/Story processing (LS), and Speech Fluency (SF) — with good fit in both unregressed and residualized score (Woods et al., 2026).

The present paper applies that bifactor decomposition to patterns of individual cognitive variation in the CCAB normative sample, focusing on the age ≥ 65 subsample (N = 620) where Mild Cognitive Impairment (MCI) is of central clinical concern. We pursue three aims. First, we quantify the proportion of CCAB domain-composite variance attributable to g across age strata spanning the full adult lifespan (18–89), testing the prediction that dedifferentiation should manifest as a monotonic increase in g share with age using both unregressed and demographically residualized scores. Second, we compare bifactor specific-factor scores against demographically residualized domain composites in convergent ranking, in variance decomposition, and in MCI detection in the lower tail of the distribution. Third, we characterize the distribution of profile patterns in the age ≥ 65 subset using an exploratory threshold-based grouping to compare the detectability of domain-specific cognitive deficits.

Methods

Participants

We analyzed the full CCAB normative sample (N = 1,916) described in detail in Woods et al. (2026). Participants were native English-speaking adults recruited from multiple cohorts contributing to the CCAB normative database. The age ≥ 65 subset (n = 620) had mean age 71.9 years (SD = 5.2), with female-majority (58.2%) and racial/ethnic diversity broadly comparable to the full sample. Inclusion criteria required native English language proficiency, adequate visual and auditory acuity for telemedical assessment, sufficient computer-use proficiency to operate the testing interface, and absence of self-reported neurological diagnosis (history of stroke, traumatic brain injury, or dementia). Participants provided written informed consent under WIRB protocol 20201196.

Bifactor confirmatory factor analytic model

The bifactor confirmatory factor analytic model was estimated on the demographically unresidualized and residualized indicators from the full normative sample using the lavaan package (Rosseel, 2012) in R version 4.4.0 with maximum likelihood and Satorra-Bentler robust (MLR) estimation. The 96 test scores were grouped into 5 sets of paired indicators after grouping based on correlation analysis (Woods et al, 2026). Both models specified a general cognitive ability factor (g) loading on all ten indicators, plus four orthogonal specific factors (PS, LS, EM, SF) loading on their corresponding parcel pairs. Executive Function indicators loaded exclusively on g, with no specific EF factor estimated. Model fit was acceptable for unresidualized scores (robust CFI = 0.976, robust RMSEA = 0.076, SRMR = 0.027) and improved for residualized scores (robust CFI = 0.982, robust RMSEA = 0.050, SRMR = 0.024). Full model specification, fit indices, and replication results are reported in the companion manuscript.

Individual-level factor scores and domain composites

Five bifactor factor scores (g, EM, PS, LS, SF) were extracted for the residualized indicators for each participant using the empirical Bayes “regression” method in lavaan, with EF loading exclusively on g. In parallel, demographically residualized cluster scores were computed as the mean of C-model residualized indicator scores within each domain (EF, EM, LS, PS, SF), yielding five domain composites (DOM_EF, DOM_MEM, DOM_LS, DOM_PS, DOM_SF). The domain composites and bifactor factor scores are matched in dimensionality (five each) but differ in construction: the composites average across indicator-level residuals with no factor model imposed, while the factor scores partition variance into a general factor (g) and four orthogonal specific factors.

Profile group classification

Profile group membership in the age ≥ 65 subset was characterized using an exploratory bottom-7% threshold applied independently to each of the five residualized factor scores. The 7% rate was chosen to approximate the standard definition (z-score < -1.5) used to define MCI (Roberts & Knopman, 2013). Because the threshold was applied to each of five correlated factors independently, resulting aggregate classification rates will over-identify MCI. Participants were partitioned into mutually exclusive groups: Normal (no score below threshold); Isolated g (g below threshold, all four specific factors above); Isolated PS/LS/EM/SF (a single specific factor below threshold, g and the other three above); g+ (g below threshold together with one or more specific factors); and g−, multi-factor (g above threshold with two or more specific factors below). Thresholds were computed on the full normative sample (N = 1,916) and applied to the age ≥ 65 subset.

Statistical analysis

Two statistical approaches were used to address reliability concerns inherent to a data-driven subgroup analysis with small within-group sample sizes. First, 2,000-iteration bootstrap resampling estimated uncertainty in subgroup-level statistics (subgroup sizes, factor-score means, demographic contrasts). Second, demographic contrasts against the Normal reference group were summarized as P(same sign), the proportion of 2,000 bootstrap samples in which the observed contrast preserved its direction. A pre-specified robustness criterion required bootstrap P(same sign) ≥ 95% with absolute effect size |d| > 0.3 for a contrast to be reported as robust. All analyses were performed in R version 4.4.0.

Domain composite vs bifactor factor score comparison

Domain composites were compared with bifactor factor scores at three levels. First, convergent correlations were computed between each demographically residualized domain composite (the mean of the two corresponding cluster scores) and its matching bifactor factor score. Second, the proportion of domain-composite variance attributable to g, to the domain-specific factor, and to unique variance was computed from the standardized bifactor loadings. Because g and the specific factors are orthogonal by construction in the bifactor model, these three components partition the composite variance exactly (summing to one), avoiding the over-counting that arises when variance shares are estimated from correlations among non-orthogonal factor-score estimates.

The primary significance test of the dedifferentiation hypothesis used the bifactor g-loadings directly. Each subject's ten demographically residualized cluster parcel scores and bifactor g factor score were used to compute within-bin Pearson correlations corr, which serve as empirical g-loadings under the assumption that the bifactor model fits each stratum reasonably well. Six non-overlapping age bins were defined: a <50 reference group (N = 742), then 50–59 (N = 359), 60–64 (N = 194), 65–69 (N = 233), 70–74 (N = 199), and 75+ (N = 187). Each older bin was compared individually to the <50 reference using a two-sided Fisher z-difference test on each parcel's correlation, with z-difference standard error $\sqrt{[1/(N\_bin-3) + 1/(N\_ref-3)]}$; domain-level loading comparisons used the mean of the two parcels per domain with a pooled standard error.

To test whether the magnitude of dedifferentiation differed between the unregressed and demographically residualized scorings, we computed, for each domain, the change in g-loading from the <50 to the 75+ bin separately for each scoring (ΔA for unregressed: the unregressed domain composite correlated with the unregressed factor score; ΔB for residualized: the residualized composite correlated with the residualized score) and tested their difference (ΔA − ΔB) by nonparametric bootstrap. Because the two scorings are computed on the same individuals within each bin (rendering the loadings dependent) but on different individuals across bins, the bootstrap resampled participants independently within each age bin (5,000 replications), recomputing all four correlations and the difference-of-differences on each replication; two-sided p-values were the proportion of replications falling on the opposite side of zero from the point estimate, with 95% confidence intervals from the 2.5th and 97.5th percentiles.

Results

Cognitive dedifferentiation gradient across age strata

Bifactor g-loadings were examined across six non-overlapping age bins (Table 1, Figure 1): a young-adult reference group under age 50 (N = 743, mean age 34), age 50–59 (N = 359, mean age 55), age 60–64 (N = 194, mean age 62), age 65–69 (N = 233, mean age 67), age 70–74 (N = 200, mean age 72), and age 75 or older (N = 187, mean age 79). The empirical g-loading for each cluster parcel within each bin was computed as the correlation between that parcel's residualized score and the bifactor factor score for both unregressed and residualized scores. Each older bin was compared individually against the <50 reference using two-sided Fisher z-difference tests on the parcel-level loadings, with domain-level loadings obtained by averaging the two parcels per domain. The dedifferentiation hypothesis predicts that g-loadings on each parcel should increase progressively with age with the general factor accounting for a progressively larger share of variance in domain-relevant test performance.

Table 1. Cognitive dedifferentiation: bifactor g-loadings across non-overlapping age bins, for unregressed and demographically residualized factor scores, with significance tests against the <50 reference group.

Domain	<50 (ref)	50–59	60–64	65–69	70–74	75+
N	*743*	*359*	*194*	*233*	*200*	*187*
Panel A. Unregressed g-scores						
EF	0.96	0.96	0.97	0.97**	0.97*	0.98***
LS	0.72	0.72	0.79*	0.75	0.75	0.80*
EM	0.57	0.55	0.70**	0.66	0.72**	0.73***
PS	0.42	0.46	0.64***	0.58**	0.49	0.62***
SF	0.49	0.54	0.71***	0.65**	0.63**	0.70***
Panel B. Residualized g-scores (C-model)						

Domain	<50 (ref)	50–59	60–64	65–69	70–74	75+
EF	0.98	0.98	0.98	0.98	0.98	0.99**
LS	0.40	0.47	0.38	0.45	0.50	0.65***
EM	0.37	0.42	0.42	0.40	0.58***	0.62***
PS	0.28	0.36	0.39	0.44*	0.32	0.44*
SF	0.39	0.46	0.53*	0.50	0.43	0.56**

Note. Values are within-bin mean g-loadings, computed as the correlation between each domain's raw cluster parcel score (mean of two parcels per domain) and the bifactor g-factor score from the corresponding solution (Panel A: unregressed g; Panel B: C-model residualized g; N = 1,916). Significance is from two-sided Fisher z-difference tests against the <50 reference: * p < 0.05, ** p < 0.01, *** p < 0.001. Panel A correlates each unregressed domain-cluster composite (mean of the two cluster parcels) with the unregressed g factor score; Panel B correlates each residualized composite (mean of the two cluster parcels) with the residualized g factor score. In residualized scores (Panel B), LS and EM show the steepest age-related increases (LS: 0.39 → 0.64, +64%; EM: 0.36 → 0.59, +64%, both p < 0.001 at 75+); PS and SF show smaller increases (PS: 0.28 → 0.44, +57%; SF: 0.39 → 0.56, +44%). EF is near-maximally g-saturated by construction (g-only loading) and is approximately 0.98 across all strata in both solutions.

Interpretation

Both panels show the same qualitative pattern—rising g-loadings with age—confirming dedifferentiation in both unregressed and demographically residualized scores. In the residualized analysis (Panel B), which indexes dedifferentiation net of demographic confounding, LS and EM show the steepest gradients (both +64% from <50 to 75+) and PS and SF smaller but reliable increases (+57% and +44%). That the gradient persists after the C-model has removed all significant demographic predictors—including age —indicates that the dedifferentiation is not simply a consequence of the age influence on test scores. The unregressed panel (A) shows larger absolute g-loadings throughout, because unregressed composites retain age-related variance shared with g; the residualized panel isolates dedifferentiation in the absence of age and other demographic contributions.

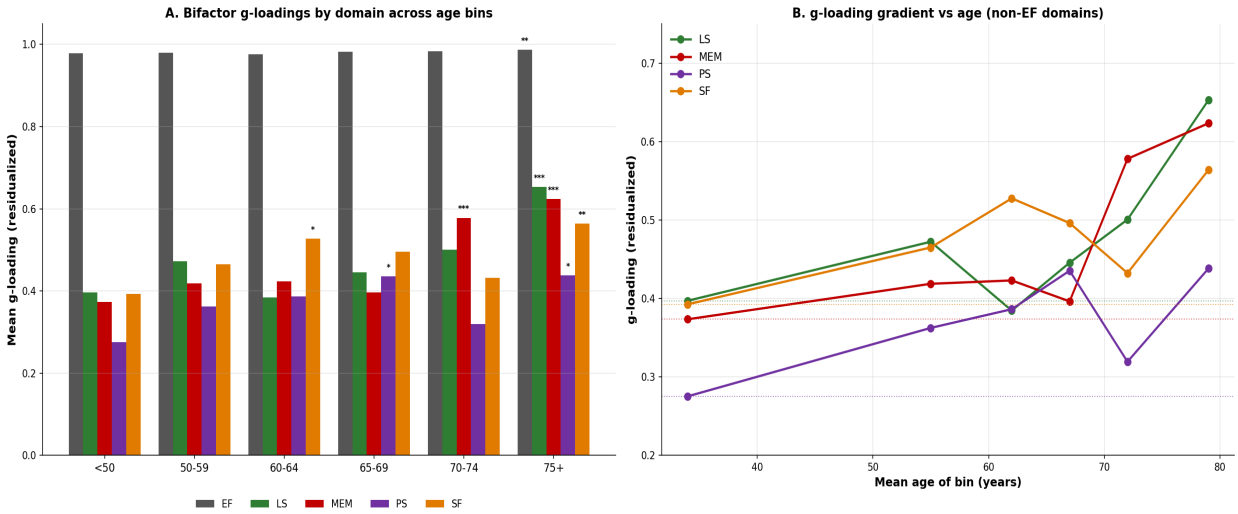

Figure 1. *Age-stratified bifactor g-loadings across non-overlapping 5- to 10-year age bins, comparing each older bin to a reference <50 group for residualized scoresin each of 5 domains. Panel A: mean g-loading per domain (mean of two cluster parcels per domain) across six age bins: <50 (N = 743, reference), 50–59 (N = 359), 60–64 (N = 194), 65–69 (N = 233), 70–74 (N = 200), and 75+ (N = 187). Significance markers above each older bin show p-values from two-sided Fisher z-difference tests against the <50 reference: * p < 0.05, ** p < 0.01, *** p < 0.001. Panel B: g-loading gradient plotted against the mean age of each bin for the four non-EF domains, with the <50 reference loading shown as a dotted horizontal line for each domain. LS and EM show the steepest gradients, with LS rising from 0.39 in <50 to 0.64 in 75+ (a 64% increase, p < 0.001). Empirical g-loadings are computed as the within-bin correlation between each domain composite (mean of the two cluster parcels) and the bifactor g factor score from the corresponding solution derived from the full-sample bifactor CFA.*

Figure 1 shows dedifferentiation in from factor analysis of residualized scores. For the LS domain, the mean residualized g-loading rose from 0.39 in the <50 reference to 0.64 in the 75+ bin — a 64% increase, significant at p < 0.001 (two-sided Fisher z-difference test). EM showed a similar gradient (0.36 → 0.59, +64%, p < 0.001

at age 75+). The increase was concentrated in the oldest bins, indicating that dedifferentiation of the verbal and memory domains is most pronounced in late life. PS and SF showed smaller gradients that reached significance at age 75+ (PS: $p < 0.05$; SF: $p < 0.01$). EF g-loadings remained near ceiling across all strata (≈ 0.98), as expected from the bifactor specification in which EF parcels load exclusively on g.

As a converging confirmation using a different summary metric, the r^2 between each demographically residualized domain composite and g rose monotonically across cumulative age strata: for LS from $r^2 = 0.18$ in adults under 60 to $r^2 = 0.43$ in age ≥ 75 (a 139% increase); for EM, 0.15 → 0.39 (+160%); for PS, 0.09 → 0.19 (+111%); and for SF, 0.17 → 0.32 (+88%). The two analytic approaches — direct g-loading comparison and r^2 decomposition — converge on the same conclusion: as participants age, the general factor accounts for an increasing share of variance in domain-relevant test performance, and the domain composite progressively conflates domain-specific operations with general cognitive ability.

The differential dedifferentiation pattern across domains is informative. LS and EM, the two memory/language constructs in the bifactor solution, showed the steepest gradients with PS and SF showing smaller increases. This is consistent with two empirical observations in the cognitive aging literature. First, crystallized verbal knowledge (LS) holds up well into middle age, then begins to decline in late life; as it declines, it converges with episodic memory (EM) and with g, because shared neurodegenerative and vascular pathology affects both. Second, PS and SF have age-specific decline mechanisms (white-matter integrity for PS; motor-speech and articulatory rate changes for SF) that produce decline trajectories not driven by shared cognitive deterioration; their specific factors retain dominance across the lifespan. Importantly for clinical application, the rapid rise of g share in the LS and EM composites means that in older adults the bifactor specific factors LS and EM isolate substantially more domain-specific signal than the corresponding domain composites. In the age ≥ 75 stratum, a participant's LS domain composite is 42% general ability and 58% domain-specific; the LS factor score isolates the domain-specific component.

Domain composites vs bifactor factor scores

Domain composites and bifactor factor scores were compared across the full sample (Figure 2). Convergent correlations between each demographically residualized domain composite and its matching bifactor factor score were uniformly high: DOM_EF vs g $r = 0.98$, DOM_LS vs LS $r = 0.93$, DOM_MEM vs MEM $r = 0.94$, DOM_PS vs PS $r = 0.95$, and DOM_SF vs SF $r = 0.93$. The two scoring approaches therefore rank participants similarly on each dimension. The 0.98 correlation between DOM_EF and g operationalizes at the individual level the parent finding that EF parcels loaded exclusively on g in the bifactor solution: the EF domain composite and the general factor score were functionally interchangeable.

The variance decomposition (Figure 2, panel B) clarifies where the two scoring approaches differ. Decomposing each composite's variance using the orthogonal standardized bifactor loadings (so that the g, specific, and unique components sum to one), the non-EF domain composites carry only 9–16% of their variance from g and 51–73% from the domain-specific factor, with the remainder unique: DOM_LS is 15% g, 73% specific, 11% unique; DOM_MEM is 14% g, 51% specific, 35% unique; DOM_PS is 9% g, 70% specific, 21% unique; DOM_SF is 16% g, 58% specific, 26% unique. DOM_EF is 76% g with no specific factor and 24% unique variance, consistent with the bifactor specification in which EF parcels load on g alone. The bifactor specifics, being orthogonal to g by construction, isolate the domain-specific variance that the domain composites carry alongside their g component. Off-diagonal correlations confirm this discriminant separation: DOM_PS correlates 0.48 with SF, DOM_LS correlates 0.44 with g and 0.34 with MEM, and DOM_SF correlates 0.45 with g and 0.42 with PS — cross-loadings that the bifactor decomposition removes.

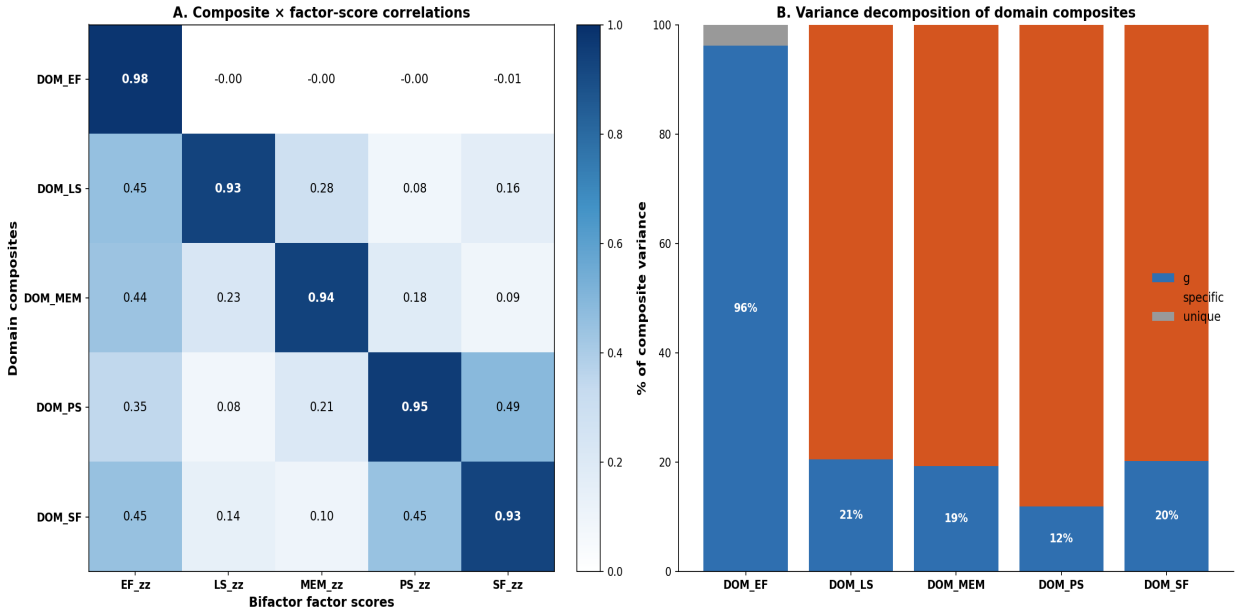

Figure 2. *Head-to-head comparison of demographically residualized domain composites (mean of two cluster scores) and bifactor factor scores in the full sample (N = 1,916). Panel A: 5×5 correlation matrix between domain composites (rows) and bifactor factor scores (columns); convergent correlations on the diagonal (r = 0.93–0.98) show that the two scoring approaches rank participants similarly on each dimension. Panel B: variance decomposition of each domain composite into the proportion explained by g (blue), by the matching specific factor (orange), and unique residual (gray). Non-EF domain composites are approximately 12–21% g and 86–91% specific in the full sample; DOM_EF is 0.96 g.*

Selective-deficit detection differed significantly between the two scoring approaches. Applying the bottom-7% threshold independently to each scoring scheme in the full sample (with the additional requirement that all other dimensions remain above the median), the bifactor LS flagged 18 participants with selective LS deficits, compared to only 3 flagged by the DOM_LS composite — a six-fold difference. The pattern was preserved in age ≥ 65 (6 FA vs 1 DOM). Cohen's kappa between the two approaches was modest (κ = 0.28 for LS, κ = 0.28 for EM, κ = 0.00 for PS, κ = 0.50 for SF in the full sample), indicating that despite the near-equivalence in convergent ranking, the detection of selective deficits diverged substantially. In participants with preserved g, the domain composite was pulled toward the population mean by the preserved general ability (g-contribution), masking the domain-specific deficit. The bifactor decomposition, by partialling g out of the specific factor, exposes the deficit.

Profile group composition in the age ≥ 65 subsample

Among 620 participants aged ≥ 65, 471 (76.0%) were classified as Normal, and 149 (24.0%) fell below the bottom-7% threshold on g or at least one specific factor and were partitioned into seven mutually exclusive profile groups (Table 2). Thresholds were applied independently to g and four correlated specific factors, so the aggregate classification rate would over diagnose clinical MCI. Across the single-domain Isolated groups,75–91% of factor-flagged cases were also flagged by the matching residualized domain composite (Table 2, final column), meaning that 9–25% of selective deficits would be missed by domain-composite scoring. Critically, this detection gain is concentrated where the dedifferentiation mechanism predicts it should be — in individuals whose general ability is preserved. (equivalent to a z-score < -1.5)

Table 2. *Profile group distribution in the age ≥ 65 subset (N = 620). The final column gives the percentage of each group also flagged (bottom 7%) on the matching residualized domain cluster composite; for Isolated g the executive function (EF) composite is used as the observed-score proxy for g, because EF parcels load almost exclusively on the general factor, and for the multi-factor groups (g+, g−) it reports the percentage flagged on at least one of the four specific-domain composites. The complement (100% minus the value) is the percentage of factor-flagged selective deficits that the domain composite fails to detect — 9–25% across the single-domain Isolated groups — quantifying the additional selective-deficit sensitivity of bifactor factor scores.*

Profile group	N	% of N = 620	% also flagged by domain composite
Normal	471	76.0%	—
Isolated g	17	2.7%	47.1%
Isolated PS	19	3.1%	68.4%
Isolated LS	24	3.9%	54.2%
Isolated EM	21	3.4%	52.4%
Isolated SF	18	2.9%	77.8%
g+ (g and ≥1 factor)	27	4.4%	77.8%
g−, multi-factor	23	3.7%	91.3%

Each profile group shows a characteristic domain-score signature (Table 3). In each Isolated subgroup, the affected domain shows a deep mean deficit (−1.22 to −1.60 in the residualized domain cluster score) while unaffected domains remain within approximately 0.25 SD of zero. The Isolated g subgroup shows the inverse pattern: a deep deficit on EF (−1.36, reflecting EF's near-exclusive g loading) with the four specific domains only mildly depressed. The g+ subgroup is depressed across all domains (−0.90 to −1.38), and the g−, multi-factor subgroup shows multiple moderate domain deficits (e.g., PS −1.13, EM −0.64) with EF relatively preserved (+0.10). These signatures were stable across bootstrap resampling.

Table 3. Mean residualized domain cluster scores by profile group, age ≥ 65.

Profile group	EF	LS	EM	PS	SF
Normal	0.04	0.17	0.10	0.09	0.18
Isolated g	−1.48	−0.63	−0.85	−0.23	−0.47
Isolated PS	0.05	0.14	−0.18	−1.17	−0.22
Isolated LS	0.13	−1.20	0.08	0.17	0.02
Isolated EM	0.18	−0.19	−1.12	0.10	0.00
Isolated SF	−0.08	−0.21	0.15	−0.14	−1.36
g+ (g and ≥1 factor)	−1.34	−1.26	−0.89	−0.79	−0.95
g−, multi-factor	0.04	−0.45	−0.70	−1.12	−0.75

Figure 3 shows the incidence of domain-specific abnormalities for different g-factor quintiles identified in domain composites and latent factor scores. In the lowest-g quintiles the domain composites flagged more individuals than factor scores, because low general ability depressed the domain-composite scores more than factor scores. In the highest-g quintiles the pattern reversed, with factor scores continuing to detect selective domain deficits (10% in Q5) that composite scores almost entirely miss (1% in Q5) because of the uncorrected influence of high g scores. The crossover demonstrates that the bifactor decomposition isolates domain-

specific signal from general ability, recovering selective deficits in high-functioning older adults that composite scoring obscures.

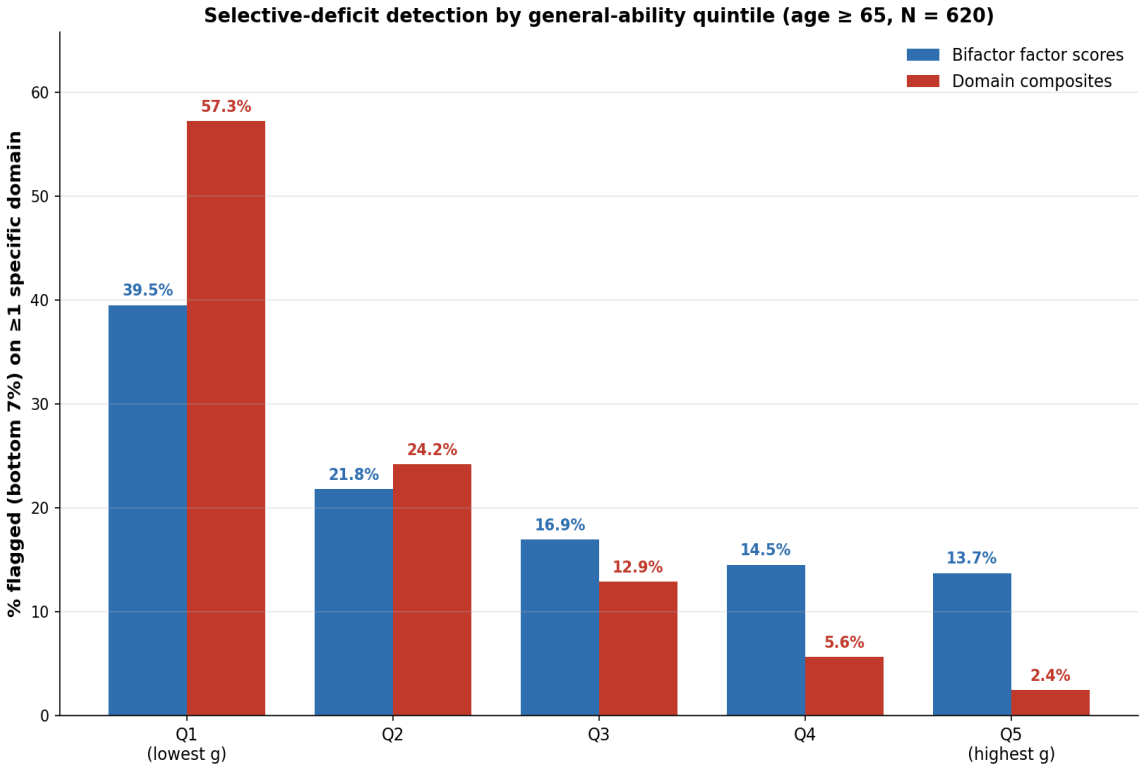

Figure 3. *Selective-deficit detection as a function of general-ability level, age ≥ 65 (N = 620). Participants are binned into quintiles of the residualized general factor (g), from lowest (Q1) to highest (Q5) general ability. The y-axis shows the percentage of each quintile flagged in the bottom 7% on at least one of the four specific domains, scored two ways: bifactor factor scores (blue) and residualized domain cluster composites (red); the general factor itself is excluded from the count. . Each quintile n ≈ 124.*

Concordance of factor-based and composite-based detection

To quantify how often the two scoring approaches identify the same individuals, we compared the bottom-7% flags of each latent factor score with those of its matching domain composite in the age ≥ 65 sample (Figure 4). Because each score flags exactly 44 participants (the bottom 7% of 620), the informative quantity is the overlap between the two flagged sets. Agreement was highest for the general-ability contrast—the g factor score versus the EF domain composite—at 69% (Cohen's κ = 0.80), consistent with EF's near-exclusive loading on g, which renders the EF composite an effective g proxy. Agreement was uniformly lower for the specific domains and declined as the specific factor became less g-saturated: LS 66% (κ = 0.78), PS 63% (κ = 0.76), SF 57% (κ = 0.71), and EM 54% (κ = 0.68). For every specific domain, roughly one-third to one-half of the individuals flagged by one method were not flagged by the other.

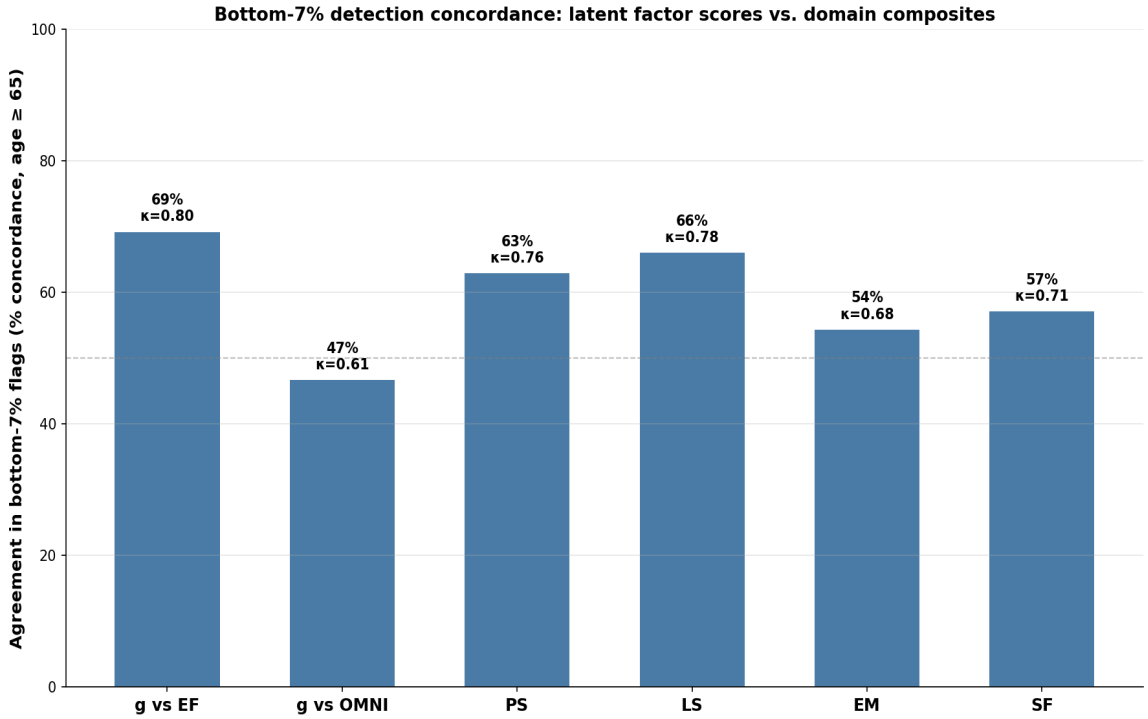

Figure 4. *Bottom-7% detection concordance between latent bifactor scores and domain composites, age ≥ 65 (N = 620). Each bar gives the percentage agreement (Jaccard: intersection ÷ union of the two flagged sets) between the factor-score–defined bottom 7% and the composite-defined bottom 7%, with Cohen's κ above each bar. The general-ability contrast pairs the g factor score with the EF domain composite (EF loads near-exclusively on g); each specific-domain contrast pairs that domain's latent factor score with its matching two-cluster residualized composite. Each score flags exactly 44 participants, so the informative quantity is how often the two methods flag the same individuals. Dashed line marks 50% agreement. A second general-ability contrast pairs the g factor score with an OMNI composite (the mean of all five domain composites); its lower agreement (47%, κ = 0.61) shows that averaging in the orthogonal specific-factor variance dilutes the g signal relative to the EF composite.*

Pattern correspondences to recognized cognitive syndromes

The profile patterns identified through bifactor decomposition invite comparison with cognitive syndromes described in the broader neuropsychological and aging literature. Isolated memory (EM) deficits described in preclinical Alzheimer's disease (Vannini et al., 2017). The Isolated g profile (reduced general ability with the four specific domains relatively preserved) may warrant comparison with non-amnestic executive-function MCI variants and with broadly lower cognitive baselines associated with educational and social determinants of cognitive function. The g+ profile (g deficit with one or more specific deficits) and the g−, multi-factor profile (multiple specific deficits with preserved g) resemble multi-domain MCI phenotypes. The Isolated LS and Isolated SF profiles invite comparison with the lexical-semantic and motor-speech disturbances established in primary progressive aphasia variants (Mesulam et al., 2014; Hodges & Patterson, 2007).

Discussion

We applied a bifactor confirmatory factor decomposition of comprehensive computerized neuropsychological data to general patterns of individual cognitive variation in the CCAB normative database, with direct comparisons against demographically residualized domain composites. The central empirical finding was a dedifferentiation gradient: bifactor g-loadings cluster parcels rose substantially with age. The increase was concentrated in the oldest bins and most pronounced after age 70. A converging analysis using cumulative age strata showed that the proportion of domain-composite variance attributable to g rose monotonically with age, from r^2 = 0.09–0.17 for non-EF domains in young adults to r^2 = 0.20–0.41 in adults aged 75 and older. This pattern is the predicted signature of age-related dedifferentiation (Blum & Holling, 2017; Breit et al., 2022; Ghisletta & Lindenberger, 2003; Tucker-Drob, 2009; Koen & Rugg, 2019) and has direct clinical consequences. As participants age, domain composites become progressively more contaminated by general cognitive ability. Consequently, the bifactor decomposition does proportionally more work in older participants by partialling out the g component that increasingly dominates composite scoring. Profile-group analysis in the

age ≥ 65 subsample partitioned participants into mutually exclusive groups whose distribution and demographic correlates are consistent with established syndromes of age-related cognitive decline. The fact that 78–84% of selective specific-factor deficits in the full sample occurred in participants with intact general cognitive ability makes bifactor decomposition clinically useful for early detection of age-related cognitive change.

Dedifferentiation and the case for bifactor decomposition in older adults

The most consequential empirical finding of the present work is the dedifferentiation gradient in bifactor g-loadings. Across the four non-EF domains, the within-bin g-loading rose substantially with age. The gradient was steepest for LS ($p < 0.001$) and EM domains ($p < 0.001$), but was also significant for PS ($p < 0.05$) and SF ($p < 0.01$). The differential rate at which loadings rise across domains — fastest for LS and EM, slower for PS and SF — was not an artifact of the model: g-loadings increased even though the bifactor model itself is fit on the full sample with a fixed set of parameters, so the increase reflects empirical within-bin covariance structure. A converging analysis confirmed this pattern at the variance-decomposition level: the proportion of variance in each demographically residualized domain composite attributable to g rose monotonically across cumulative age strata. The pattern is what age-related dedifferentiation predicts (Spearman, 1927; Blum & Holling, 2017; Breit et al., 2022) and is supported by a large empirical and theoretical literature. Ghisletta and Lindenberger (2003) demonstrated direct longitudinal evidence for ability dedifferentiation in old age in the Berlin Aging Study. Tucker-Drob (2009) applied nonlinear factor analytic models to 6,273 individuals and documented graded ability differentiation across the lifespan. At the neural level, Park et al. (2004) showed reduced functional specialization of ventral visual cortex in older adults, and the broader neural-dedifferentiation framework reviewed by Koen and Rugg (2019) links these representational changes to age-related cognitive decline.

A methodological feature of the present analysis distinguishes it from this prior literature: we examined dedifferentiation in both raw (unregressed) and demographically regressed indicators. Previous tests of age dedifferentiation have been conducted on raw or sample-standardized indicators (e.g., Tucker-Drob, 2009; Ghisletta & Lindenberger, 2003). In parallel, the clinical neuropsychology literature has developed demographically corrected normative scores that regress each test on age, education, and sex, but applies the correction either to support individual-level interpretation or to the derived factor scores after the factor model has been estimated on uncorrected data (e.g., Kiselica, Webber, & Benge, 2020). Here, the comprehensive C-model correction removed age, education, sex, vocabulary, race/ethnicity, and the other retained demographic predictors from every measure and then asked whether the age-related increase in g-loadings persisted in the corrected metric.

This design tests a more conservative question than the classical one: not merely whether cognitive abilities become more intercorrelated with age (which a shared age gradient alone would produce), but whether a residual structural convergence remains after the demographic gradient — including age — has been removed from the indicators themselves. It might be objected that residualizing the indicators on age removes the very variance that dedifferentiation addresses. The two analyses should therefore be read as complementary bounds: the unregressed solution, in which g-loadings rose more steeply and earlier, gives an upper-bound estimate of age dedifferentiation that includes the shared age gradient, whereas the demographically corrected solution gives a conservative lower-bound estimate net of the direct impact of age on indicators. That a reliable gradient survives demographic correction indicates that age-related dedifferentiation in these domains is not wholly an artifact of correlated demographic decline, while the larger unregressed estimates indicate that a substantial portion of the conventionally measured effect is demographically mediated. The true dedifferentiation effect is bracketed between these bounds.

Clinical implications

In adults aged 75 and older, using domain composites without g-partialling means that approximately 39–43% of what is reported as “episodic memory” or “lexical/story processing” performance is in fact general cognitive ability. For a participant with preserved g, the g component of the domain composite is pulled toward the population mean by the preserved general ability, masking potential domain-specific deficits. Bifactor scoring partials g out of the specific factor by construction, exposing the deficits that domain composites conflate with g. The selective-LS detection comparison in the present analysis illustrates the magnitude of this difference: in the full sample, LS flagged 18 selective-LS cases compared to only 3 flagged by the DOM_LS composite, a six-fold difference preserved in the age ≥ 65 subsample (6 versus 1). The differential dedifferentiation pattern

across domains is also informative for clinical application. LS and EM showed the steepest gradients, consistent with the convergence of crystallized verbal knowledge and episodic memory toward shared decline as participants enter their seventies; PS and SF showed smaller gradients, consistent with their relatively more age-linear decline mechanisms (white-matter integrity for PS; motor-speech and articulatory rate for SF). In LS- and EM-focused clinical applications in older adults, bifactor specific-factor scoring is the more appropriate analytic level; in PS- and SF-focused applications, domain composites and bifactor scores carry more similar information.

The concordance analysis in Figure 4 makes this trade-off concrete at the level of individual case detection. When the clinical question concerns general cognitive status, the choice of scoring method is nearly immaterial: the g factor score and the EF domain composite agree on 69% of flagged cases ($\kappa = 0.80$), because EF loads almost exclusively on g and its composite is therefore an effective proxy for the general factor. However, for domain-selective impairment agreement falls monotonically with the g-saturation of the specific factor, reaching its minimum for Episodic Memory (54%, $\kappa = 0.68$), the least g-loaded specific factor after residualization. For each specific domain, one-third to nearly one-half of the individuals flagged by a domain composite are not the individuals flagged by the corresponding latent factor score, and vice versa. The two methods converge on who is generally declining but disagree on who has a selective deficit—precisely the discrimination that profile-based assessment is meant to provide. The practical conclusion is therefore conditional on the diagnostic question: for a global impairment screen, domain composites and factor scores are interchangeable and the simpler composite suffices; but for identifying isolated, domain-specific deficits—especially in individuals whose general ability is preserved—only the bifactor factor score isolates the specific signal from the general background, and composite-based profiling will misassign a substantial minority of cases.

The dissociation between general and specific cognitive variance is a structural feature of bifactor parameterization and would be obscured in correlated-factor models, where inter-factor correlations of 0.60–0.85 in published UDS analyses (Gaynor et al., 2025; Matusz et al., 2025) cause low specific-factor scores to track low overall scores. Clinically, this means the bifactor decomposition is the best approach to the early detection of age-related decline in older populations where dedifferentiation increases g-factor contributions to domain scores and where selective decline in one cognitive domain precedes broader cognitive deterioration.

Mechanisms of cognitive dedifferentiation

Three classes of mechanism, operating at neurochemical, structural, and network levels, are widely invoked to explain dedifferentiation. At the neurochemical level, the neuromodulation-to-representation framework of Li, Lindenberger, and Sikström (2001) proposes that age-related decline in dopaminergic neuromodulation, with parallel declines in noradrenergic and cholinergic systems, reduces neural gain in cortical circuits. Because dopaminergic and other neuromodulatory inputs sharpen the contrast between task-relevant and task-irrelevant cortical activity at the level of pyramidal-neuron gain, reduced neuromodulation produces less distinctive cortical representations: patterns of activation that were orthogonal across stimulus categories and cognitive operations in young brains overlap more, and are harder to distinguish in old brains. Behaviorally, this representational smearing appears as increased between-task correlations, which factor-analytically isolates as an inflated general factor and reduced specific-factor variance — the dedifferentiation signature.

At a structural level, Lindenberger and Ghisletta (2009) provided an alternative explanation. Declines across cognitive abilities, sensory functions, and motor functions reflect brain-wide changes in white-matter integrity (e.g., fractional-anisotropy losses in long-range frontotemporal and frontoparietal tracts), an exponential increase in periventricular and deep white-matter hyperintensities reflecting small-vessel cerebrovascular disease, cortical thinning (steeper in association cortex than in primary sensory cortex), and medial temporal lobe atrophy disproportionately affecting hippocampal subfields and entorhinal cortex. These produce a shared-decline signal that increases the g-saturation of domain composites in older adults.

At a network level: resting-state fMRI demonstrates that large-scale brain networks (default mode, frontoparietal control, dorsal attention, salience) lose their functional segregation with age — within-network connectivity weakens and between-network connectivity strengthens, yielding a more uniform, less modular brain lacking fully distinct circuits for distinct cognitive operations. Park et al. (2004) provided the foundational demonstration that ventral visual cortex shows reduced category-selective specialization in older adults, and

the broader neural-dedifferentiation literature reviewed by Koen and Rugg (2019) extends the principle across sensory, motor, and prefrontal regions.

The differential gradient across domains in the present data — LS and EM dedifferentiating steeply, PS and SF more slowly — maps onto these mechanisms in an informative way. LS (lexical/story processing) and EM (memory) share dependence on temporal-lobe and frontotemporal-network integrity: hippocampal-entorhinal-perirhinal circuitry for episodic memory, anterior and lateral temporal cortex for lexical-semantic representation, and the frontotemporal language network for narrative comprehension and word retrieval. These substrates are jointly affected by hippocampal atrophy in normal aging, by frontotemporal white-matter degradation, and by Alzheimer-type pathology that begins in entorhinal cortex and progresses through lateral temporal regions. Because the structural and pathological mechanisms hitting LS and EM are largely shared, the two factors converge with each other and with g as participants age. PS (processing speed) and SF (speech fluency) depend on relatively more distinct neural substrates — PS on widespread white-matter integrity and subcortical-cortical loops, SF on motor-speech and articulatory circuits including premotor cortex, basal ganglia, cerebellum, and brainstem nuclei. Their decline trajectories are less synchronized with the temporal-lobe pathology that hits LS and EM together. PS and SF therefore retain more domain-specific variance into the oldest age strata.

These mechanisms also clarify why age influences the latent structure on two distinct levels, and why demographic correction separates them. Age acts first on the level of each indicator through largely peripheral or generic channels — slowed central and peripheral conduction, delayed motor activation and response execution, and reduced sensory acuity — that depress many scores in parallel without acting on the cognitive construct each test is meant to tap. Because this peripheral drag is shared across indicators, it inflates the covariance among them and, in an unresidualized solution, is absorbed as a broader general factor. Regressing each indicator on age and other demographic predicts removes this level shift and the covariance it induces,

The dedifferentiation gradient documented here operates on a different level: it is not a shift in indicator means but an age-related change in the loadings themselves — an age-by-coupling interaction — whereby a given ability draws more on domain-general resource and less on its domain-specific component in older than in younger adults. Because this interaction acts on the covariance structure rather than on indicator levels, it is invisible to mean-based demographic correction and survives in the residualized metric, where its basis is necessarily central rather than peripheral: declining domain-general neuromodulatory and executive resource, loss of network segregation, and correlated neuropathological burden. A formal comparison of the age-related rise in g-loading (<50 to 75+) between the two scorings supports this two-level account. The speed-loaded domains behave as a peripheral-channel account predicts — processing speed and speech fluency show numerically steeper dedifferentiation in the unregressed than the residualized metric. However, LS dedifferentiation is significantly steeper in the residualized metric ($\Delta = 0.25$, $p < 0.001$), with episodic memory showing a weaker version of the same pattern ($p = 0.05$). This is the signature of demographic suppression rather than inflation: because LS is a crystallized ability that improves with age, the positive demographic (age and vocabulary) trend on LS partially masks its rising g-coupling in the unregressed scores, and removing that trend unmasks the dedifferentiation. Demographic correction therefore does not uniformly attenuate dedifferentiation; it modestly attenuates it for fluid, peripherally loaded domains and amplifies it for crystallized LS, because demographic variance acts in opposite directions for the two ability types.

Domain-specific deficits in high functioning individuals

A further consequence of bifactor parameterization worth highlighting is that specific-factor scores represent deviations from each participant's own general cognitive level, not from the population mean. Because g-loadings on non-EF parcels in the residualized model are modest (0.25–0.40), the g-correction applied to specific-factor scores is moderate for most participants. However, for participants at the upper end of the g distribution, the g-correction becomes substantial. A participant with g = +2 SD and average raw performance on a specific domain would have a residualized cluster score near zero but a bifactor specific-factor score of approximately −0.7 — reflecting the bifactor's interpretation that average-level performance constitutes relative underperformance for someone of high general ability. The bifactor specific-factor score thus operationalizes a longstanding clinical intuition that selective deficits should be evaluated relative to a patient's own ability level rather than to the population mean. This is also the mechanism behind the six-fold difference in selective-LS detection between bifactor scoring and domain composites in the present analysis: in participants with

preserved g, partialling out the g component unmasks domain-specific deficits that the composite obscures. For early detection of cognitive change in high-functioning individuals whose absolute performance remains within population norms, the bifactor specific-factor score may be more sensitive. The flip side of this property is that bifactor scores are inherently relative measures: a participant's absolute level of impairment cannot be inferred from a specific-factor score alone, which is why we recommend reporting raw cluster, residualized cluster, and bifactor specific-factor scores in parallel rather than substituting one for another.

Limitations

The profile groups described here are drawn from a community-dwelling normative sample without confirmed clinical diagnoses, biomarker assessment, imaging, or longitudinal follow-up. Our central claim is therefore a measurement claim, not a diagnostic claim: cognitive dedifferentiation makes bifactor decomposition progressively more clinically valuable as participants age. In addition, the bottom-7% threshold was an exploratory approximation rather than a clinically derived cutoff; within-subgroup individual classification stability was moderate across the smaller profile groups, indicating that while group-level patterns are reproducible, individual classifications carried meaningful uncertainty. Moreover, within-subgroup sample sizes were small (n = 11 to 26 across the seven non-Normal groups), limiting confident inference about subgroup-specific characteristics.

The CCAB normative sample on which the demographic-residualization model (C-model) and bifactor parameters were trained reflects Northern California recruitment. While the sample is demographically diverse (37% White, 23% Black, 18% Asian, 22% other race/ethnicity in the full sample; 26% with post college-level educational attainment), it is concentrated in the San Francisco Bay Area and surrounding metropolitan region. The demographic-correction equations that anchor individual interpretation — and therefore the factor scores that drive the profile classifications reported above — may not generalize unchanged to other regions. Three sources of regional variation are likely to matter. First, educational attainment distributions differ substantially across U.S. regions; an eight-year-of-education participant in our Northern California sample occupies a different percentile than an eight-year-of-education participant in a different region with a different overall educational distribution. Second, the racial and ethnic composition of the Northern California normative sample (notably the high Asian representation reflecting Bay Area demographics, and the predominantly urban/suburban recruitment) does not match the composition of other U.S. regions or rural populations; the demographic-correction coefficients trained on this sample may over- or under-correct in regional applications. Third, language and bilingualism patterns differ regionally; the demographic-correction model trained on English speakers in Northern California may not extend cleanly to populations with different bilingual exposure patterns. Multi-site replication with deliberate geographic and rural representation — ideally including Southern, Midwestern, Mountain West, and rural sites alongside additional urban regions — is required before the bifactor profile patterns reported here can be confidently applied clinically beyond the geographic region in which the normative sample was acquired.

Conclusion

Cognitive dedifferentiation makes bifactor decomposition progressively more clinically valuable as participants age: the proportion of CCAB domain-composite variance attributable to g rises from approximately 16% in young adults to 35–42% in adults aged 75 and older for memory and lexical/story processing, and enables bifactor specific-factor scores to isolate domain-specific signals that domain composites conflate with general ability.

Declarations

Funding. This work was supported by the National Institute on Aging (R44AG080951, R44AG097322). The funder had no role in study design, data analysis, data interpretation, or manuscript preparation.

Competing interests. All authors are employees of Neurobehavioral Systems, Inc., which develops the California Cognitive Assessment Battery.

Ethics approval. Participants provided informed consent under protocols approved by the Western Institutional Review Board (WIRB protocol 20201196).

Clinical trial registration. ClinicalTrials.gov registration: NCT04800588.

Data availability. Data are available for reasonable requests to the corresponding author.

Author contributions. All authors contributed to study conception, data analysis, manuscript drafting, and final approval.

References

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., Nation, D. A., Libon, D. J., Au, R., Galasko, D., & Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease, 42*(1), 275–289. https://doi.org/10.3233/JAD-140276

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns: A meta-analysis. *Intelligence, 65*, 60–66. https://doi.org/10.1016/j.intell.2017.07.004

Breit, M., Brunner, M., Molenaar, D., & Preckel, F. (2022). Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin, 148*(7–8), 518–554. https://doi.org/10.1037/bul0000379

Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., Jones, R. N., Mukherjee, S., Curtis, S. M., Harvey, D., Weiner, M., & Mungas, D. (2012). Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior, 6*(4), 502–516. https://doi.org/10.1007/s11682-012-9186-z

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., Weiner, M., & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurology, 71*(8), 961–970. https://doi.org/10.1001/jamaneurol.2014.803

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Gaynor, L. S., Lopez, F. V., Van Hulle, C. A., Li, C., Vasunilashorn, S. M., Andrews, S. J., Simone, S. M., & Mungas, D. M. (2025). Measurement equivalence of the UDS version 2.0 and 3.0 neuropsychological batteries. *Alzheimer's & Dementia, 21*(9), e70720. https://doi.org/10.1002/alz.70720

Gibbons, L. E., Carle, A. C., Mackin, R. S., Harvey, D., Mukherjee, S., Insel, P., Curtis, S. M., Mungas, D., & Crane, P. K. (2012). A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging and Behavior, 6*(4), 517–527. https://doi.org/10.1007/s11682-012-9176-1

Ghisletta, P., & Lindenberger, U. (2003). Age-based structural dynamics between perceptual speed and knowledge in the Berlin Aging Study: Direct evidence for ability dedifferentiation in old age. *Psychology and Aging, 18*(4), 696–713. https://doi.org/10.1037/0882-7974.18.4.696

Hülür, G., Ram, N., Willis, S. L., Schaie, K. W., & Gerstorf, D. (2015). Cognitive dedifferentiation with increasing age and proximity of death: Within-person evidence from the Seattle Longitudinal Study. *Psychology and Aging, 30*(2), 311–323. https://doi.org/10.1037/a0039260

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., Beaumont, J., Casaletto, K. B., Conway, K., Slotkin, J., & Gershon, R. (2014). Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society, 20*(6), 588–598. https://doi.org/10.1017/S1355617714000241

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: A unique clinicopathological syndrome. *The Lancet Neurology, 6*(11), 1004–1014. https://doi.org/10.1016/S1474-4422(07)70266-1

Kiselica, A. M., Webber, T. A., & Benge, J. F. (2020). The Uniform Dataset 3.0 neuropsychological battery: Factor structure, invariance testing, and demographically adjusted factor score calculation. *Journal of the International Neuropsychological Society, 26*(6), 576–586. https://doi.org/10.1017/S135561772000003X

Koen, J. D., & Rugg, M. D. (2019). Neural dedifferentiation in the aging brain. *Trends in Cognitive Sciences, 23*(7), 547–559. https://doi.org/10.1016/j.tics.2019.04.012

Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences, 5*(11), 479–486. https://doi.org/10.1016/S1364-6613(00)01769-1

Lindenberger, U., & Ghisletta, P. (2009). Cognitive and sensory declines in old age: Gauging the evidence for a common cause. *Psychology and Aging, 24*(1), 1–16. https://doi.org/10.1037/a0014986

Matusz, E. F., Fiala, J., Kiselica, A. M., Rosselli, M., Armstrong, M. J., Holgerson, A. A., Levy, S.-A., Arias, F., Vélez-Uribe, I., Duara, R., Curiel Cid, R. E., Loewenstein, D. A., Smith, G. E., Marsiske, M., & Asken, B. M. (2025). Cognitive factor structure of the NACC UDS-3 neuropsychological battery across ethno-racial, linguistic, and cognitive status groups. *The Clinical Neuropsychologist*. Advance online publication. https://doi.org/10.1080/13854046.2025.2576154

Mesulam, M.-M., Wieneke, C., Thompson, C., Rogalski, E., & Weintraub, S. (2014). Quantitative classification of primary progressive aphasia at early and mild impairment stages. *Brain, 137*(4), 1176–1192. https://doi.org/10.1093/brain/awu024

Nelson, H. E. (1982). *National Adult Reading Test (NART): Test manual*. NFER-Nelson.

Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). Aging reduces neural specialization in ventral visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 101*(35), 13091–13095. https://doi.org/10.1073/pnas.0405148101

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children—Second Edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly, 22*(4), 511–539. https://doi.org/10.1037/1045-3830.22.4.511

Roberts, R., & Knopman, D. S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine, 29*(4), 753–772. https://doi.org/10.1016/j.cger.2013.07.003

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.

Staffaroni, A. M., Asken, B. M., Casaletto, K. B., Fonseca, C., You, M., Rosen, H. J., Boxer, A. L., Elahi, F. M., Kornak, J., Mungas, D., & Kramer, J. H. (2021). Development and validation of the Uniform Data Set (v3.0) executive function composite score (UDS3-EF). *Alzheimer's & Dementia, 17*(4), 574–583. https://doi.org/10.1002/alz.12214

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*(4), 1097–1118. https://doi.org/10.1037/a0015864

Vannini, P., Hanseeuw, B., Munro, C. E., Amariglio, R. E., Marshall, G. A., Rentz, D. M., Pascual-Leone, A., Johnson, K. A., & Sperling, R. A. (2017). Anosognosia for memory deficits in mild cognitive impairment: Insight into the neural mechanism using functional and molecular imaging. *Neuropsychologia, 99*, 343–349. https://doi.org/10.1016/j.neuropsychologia.2017.04.002

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)*. The Psychological Corporation.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31*(2), 94–113. https://doi.org/10.1177/0734282913478030

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14*(3), 144–148. https://doi.org/10.1111/j.0963-7214.2005.00356.x

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. (2026). *Novel cognitive factor structure revealed by comprehensive computerized neuropsychological assessment* [CCAB Technical Report]. Neurobehavioral Systems, Inc. www.ccabresearch.com