Journal of *Intelligence*

MDPI

*Review*

# Estimating Premorbid IQ in Culturally and Linguistically Diverse Older Adults: A Critical Narrative Review

**Philip Sayegh [1,*], Luis D. Medina [2,3], Cecilia Y. M. Poon [4,5], Paola Suarez [6] and Christopher M. Nguyen [7]**

1 Department of Psychology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA
2 Department of Psychology, University of Houston, Houston, TX 77204, USA; ldmedina2@uh.edu
3 Department of Psychiatry and Biobehavioral Sciences, Baylor College of Medicine, Houston, TX 77030, USA
4 Psychology Department, Nebraska Medicine, Omaha, NE 68198, USA; cepoon@nebraskamed.com
5 Geriatric Psychiatry Division, Department of Psychiatry, University of Nebraska Medical Center, Omaha, NE 68105, USA
6 Department of Psychiatry and Behavioral Sciences, UCLA Health, Los Angeles, CA 90095, USA; psuarez@mednet.ucla.edu
7 Department of Psychiatry and Behavioral Health, The Ohio State University College of Medicine, Columbus, OH 43210, USA; christopher.nguyen@osumc.edu
* Correspondence: psayegh@psych.ucla.edu

**Abstract**

Estimating premorbid IQ is essential for determining whether current cognitive performance reflects decline from an individual's likely level of functioning before neurological injury or disease, yet commonly used estimators may be less valid in culturally and linguistically diverse (CALD) older adults. This critical narrative review synthesizes direct, within-population validation, and contextual evidence across established premorbid-IQ methods, newer culturally grounded and informant-based approaches, and emerging computational, longitudinal, and harmonization directions. Direct evidence on diversity-related influences remains limited, whereas within-population validation is concentrated most strongly in language-specific word-reading measures; no single method has been sufficiently validated across heterogeneous CALD older adults. We therefore propose an evidence-informed structured multimethod framework organized around four clinical tasks: Select the quantitative estimator or estimators with the strongest individual applicability; Contextualize estimates using developmental, educational, linguistic, occupational, migration, and collateral information; Compare reasonably applicable estimates for convergence, divergence, and plausible sources of discrepancy; and Report the most defensible premorbid range, evidentiary weighting, and remaining uncertainty. Contextual information should inform applicability and relative weight rather than unsupported numerical corrections. The framework is intended to structure clinical reasoning while highlighting priorities for future validation research and has not itself been validated as a standardized clinical algorithm.

**Keywords:** premorbid IQ; culturally and linguistically diverse populations; cultural neuropsychology; older adults; multilingual assessment; literacy; educational quality; word reading; neuropsychological assessment; cross-cultural assessment

Check for updates

Received: 11 August 2026
Revised: 29 October 2026
Accepted: 1 November 2026
Published: 3 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Assessment of premorbid IQ is vital given the cross-sectional nature of a neuropsychological evaluation and at the core of making correct inferences about how one's brain functions at present (van den Broek, 2025). Premorbid IQ refers to an estimate of an individual's level of intellectual functioning prior to the onset of a neurological injury,

neurodegenerative disease, or other condition capable of altering cognitive performance (Bright & van der Linde, 2020; Franzen et al., 1997). Because direct measurements obtained before illness or injury are rarely available, premorbid IQ typically must be inferred retrospectively. Such estimates provide an individualized comparison point for evaluating whether current cognitive performance represents decline relative to the person's likely pre-injury or pre-disease level of ability (Lezak et al., 2012; Strauss et al., 2006).

Premorbid IQ and cognitive reserve are related but distinct constructs. Cognitive reserve refers to the adaptability of cognitive processes, including their efficiency, capacity, and flexibility, that helps explain why individuals with similar levels of brain aging, pathology, or insult may differ in cognitive and everyday functioning (Stern, 2009; Stern et al., 2020). Educational experience, premorbid IQ, occupational complexity, literacy, and cognitively stimulating activities have commonly been used as proxies for or contributors to cognitive reserve, but none is synonymous with the construct (Nogueira et al., 2022; Stern et al., 2020). This distinction is particularly important in culturally and linguistically diverse (CALD) populations because commonly used indicators of both premorbid IQ and cognitive reserve may be shaped by educational quality, literacy, language, and sociocultural opportunity (Manly et al., 2002). For example, a low score on a language-bound word-reading estimator may coexist with substantial occupational complexity, multilingual experience, adaptive functioning, or other evidence of lifetime cognitive enrichment. Such discordance should prompt consideration of whether the estimator adequately reflects the individual's linguistic and educational history rather than automatic inference of low premorbid IQ.

Cognitive test performance is usually captured in a single session, providing insights into the examinee's current status. However, referral questions are typically related to changes in cognitive status, such as declines due to an event (e.g., head injury secondary to a motor vehicle accident) or a disease process (e.g., Alzheimer's disease). Without the context of an individual's premorbid IQ, two individuals with the same score on a test could represent two distinct realities—a normal level of functioning for one and a notable decline for the other (Crawford et al., 1998). In the absence of adequate or meaningful baseline testing to which current performance can be compared, clinicians must rely on alternative methods of inference to estimate a patient's cognitive ability before the injury or onset of the relevant disease process. In this context, estimates of premorbid IQ allow clinicians to make determinations related to the presence and magnitude of cognitive decline, thereby informing diagnosis and treatment.

The older adult population in the United States (U.S.) is highly diverse with respect to race and ethnicity, language, and education. Based on U.S. Census data, approximately one-quarter of adults aged 65 and older identified as members of a racial or ethnic minority group in 2022, including 9% of Hispanic origin (of any race), 9% African American (non-Hispanic), 5% Asian American (non-Hispanic), 0.6% American Indian and Alaska Native (non-Hispanic), 0.1% Native Hawaiian/Pacific Islander (non-Hispanic), and 0.9% identifying with two or more races (Administration for Community Living [ACL], 2024). Sixteen percent of older adults spoke a language other than English at home. Educational attainment also varied substantially across groups. Although 89% of older adults overall had completed high school and approximately one-third (33%) had obtained a bachelor's degree or higher, corresponding rates differed among African American (83% and 22%), Asian (82% and 43%), and Hispanic (63% and 16%) older adults (Administration for Community Living [ACL], 2024). These demographic, linguistic, and educational differences have important implications for estimating premorbid IQ because many commonly used estimation methods rely, directly or indirectly, on factors such as education, literacy, language, and other sociodemographic characteristics.

Terminology regarding ethnicity, cultural background, and language warrants particular care. In this review, "Hispanic/Latino" is used when referring to ethnicity or origin as defined by a data source or study sample and is not treated as synonymous with Spanish-speaking. "Spanish-speaking" is used when language is the relevant characteristic, and geographic or cultural origin is specified when it materially affects interpretation. These distinctions are important because Spanish-speaking populations are themselves highly heterogeneous with respect to national origin, migration history, education, literacy, dialect, and cultural experience.

In this review, CALD is used as a pragmatic organizing term for heterogeneity in cultural background and language experience rather than as an exhaustive taxonomy of diversity. Educational quality and continuity, literacy, migration history, socioeconomic circumstances, occupational experience, acculturation, and related life experiences are considered explicitly because they may intersect with cultural and linguistic background and influence the applicability of premorbid-IQ estimators within as well as between populations. Accordingly, reference to CALD populations in this review should not be interpreted as implying that culture and language are the only relevant dimensions of heterogeneity.

Approaches to premorbid IQ estimation generally fall into several broad categories: clinical judgment, best-performance methods, demographic regression equations, measures of abilities assumed to be relatively resistant to neurological decline ("hold" measures), word-reading and related lexical measures, and combination methods that integrate test performance with demographic predictors (Bright & van der Linde, 2020; Franzen et al., 1997). More recent work has explored structured informant-based approaches and computational methods intended to improve prediction accuracy (Apolinario et al., 2013; van der Linde & Bright, 2018). Culturally grounded test-development work also offers methodological models for more context-sensitive assessment, although such measures are not themselves premorbid-IQ estimators (Jacklin et al., 2020; LoGiudice et al., 2006). Each approach has distinct advantages, assumptions, and limitations, and no single method has been established as valid across the range of cultural, linguistic, educational, and aging-related contexts considered in this review (Apolinario et al., 2013; Briceño et al., 2023; Overman et al., 2021). Section 3 summarizes these approaches before we examine the evidence most directly relevant to CALD older adults.

The contribution of the present review lies in integrating several bodies of literature that are often considered separately when estimating premorbid IQ: conventional premorbid-estimation methods, estimator validation within particular populations and languages, cultural and linguistic neuropsychology, educational quality and literacy, migration and acculturation, and emerging methodological approaches. We further distinguish empirical studies that directly examine whether diversity-related characteristics affect the performance or applicability of premorbid-IQ estimators from within-population validation studies and from contextual evidence that informs interpretation without directly validating an estimator. This synthesis is then translated into a premorbid-IQ-specific structured multimethod framework organized around four clinical tasks—Select, Contextualize, Compare, and Report—while explicitly identifying the limitations of the evidence supporting that framework.

Although several of the principles discussed in this review are relevant across adulthood, we retain an older-adult focus because neurodegenerative disease, age-related sensory changes, cohort-linked educational experiences, and the clinical need to distinguish longstanding individual differences from acquired decline make premorbid-IQ estimation particularly consequential in later life. Evidence from broader adult samples is included when it provides empirical evidence relevant to a particular premorbid estimator or necessary contextual or methodological information; such findings are not presented as direct validation in CALD older adults when the sample does not support that conclusion.

## 2. Methods

### 2.1. Literature Search Rationale and Strategy

The present article is a critical narrative review rather than a systematic or scoping review. A narrative approach was selected because the aim was to critically integrate evidence across premorbid IQ estimation, cultural and linguistic neuropsychology, educational and literacy factors, aging, and emerging methodological approaches rather than to conduct an exhaustive study-identification or quantitative meta-analytic review.

Relevant peer-reviewed literature was identified through searches of PubMed, APA PsycINFO, APA PsycArticles, and ProQuest Central, supplemented by Google Scholar and backward citation tracking of relevant empirical studies, reviews, and methodological papers. Searches focused broadly on premorbid IQ estimation; demographic, reading-based, and other performance-based estimation methods; culturally and linguistically informed neuropsychological assessment; literacy and educational quality; bilingualism, multilingualism, acculturation, and migration; cognitive reserve; sensory and motor factors in aging; and emerging longitudinal, normative, harmonization, and computational approaches. Foundational studies were retained when necessary to describe the development and psychometric basis of established premorbid estimation methods. Literature searches were last updated in September 2026.

### 2.2. Approach to Evidence Synthesis

The literature was organized thematically according to the major approaches to premorbid IQ estimation: clinical judgment, best-performance methods, demographic regression equations, "hold" measures, word-reading and related lexical measures, combination methods, culturally grounded and informant-based approaches, and emerging computational, longitudinal, and harmonization directions.

For purposes of interpretation, we distinguished three forms of evidence. Direct evidence on diversity-related influences included empirical studies that evaluated whether cultural, linguistic, educational, demographic, migration-related, or other relevant characteristics affected the performance, calibration, stability, or predictive validity of a premorbid-IQ or premorbid-cognitive-ability estimator, or that directly compared estimator performance across relevant populations, languages, or normative contexts. Within-population validation evidence included studies that established the reliability, validity, or stability of an estimator within a particular language or population without necessarily examining whether diversity-related characteristics influenced estimation. Contextual evidence included research addressing factors such as literacy, educational quality, acculturation, socioeconomic context, cross-cultural test performance, sensory functioning, or related measurement issues without directly evaluating a premorbid estimator. Contextual evidence was used to inform interpretation and clinical considerations but was not treated as direct validation of a premorbid-IQ estimator.

## 3. Approaches to Estimating Premorbid IQ

Various approaches exist for estimating premorbid IQ. However, there is currently no universally accepted gold standard given the limitations of each method. Table 1 summarizes the evidence on reliability and validity, key population limitations, and representative evidence for the established and emerging approaches discussed in this section; where an approach represents methodological infrastructure or a future research direction rather than a validated premorbid-IQ estimator, this distinction is stated explicitly. Below is a critical summary of existing approaches.

**Table 1.** Methods and Evidence Relevant to Premorbid IQ Estimation in Culturally and Linguistically Diverse Older Adults.

| Method | Evidence on Reliability | Evidence on Validity | Key Population Limitations | Representative Evidence |
| --- | --- | --- | --- | --- |
| Clinical Judgment | No single reliability coefficient applies to unstructured clinical judgment. Estimates can vary across clinicians because the demographic and historical variables considered, and the weight assigned to them, are not standardized. | Comparative work indicates that actuarial demographic estimates can correspond more closely with obtained IQ than clinicians' unstructured estimates. Clinical judgment is also vulnerable to systematic estimation error when demographic information is weighted informally. | Potential for cultural and educational bias; inconsistent weighting of age, education, occupation, language, and life history; limited reproducibility across clinicians. | (Crawford et al., 2001; Kareken, 1997; Kareken & Williams, 1994) |
| Best Performance Method | Not a fixed psychometric instrument; reproducibility depends on which scores or historical indicators are selected as the individual's "best" performance. Normal intraindividual score variability limits stability of the estimate. | Conceptually simple but can overestimate premorbid IQ because healthy individuals normally show substantial score scatter. Direct validation relevant to CALD premorbid-IQ estimation is limited. | The highest observed score may itself be influenced by language, education, cultural familiarity, or test-specific strengths; overestimation may obscure genuine decline. | (Bright & van der Linde, 2020; Franzen et al., 1997) |
| Demographic-Only Regression Equations | The same equation is reproducible when demographic variables are coded consistently, but individual prediction error can be substantial and estimates regress toward the population mean. | Can provide useful prediction independent of current neurological status, but accuracy is weaker at IQ extremes and transportability depends on the development sample. Evidence relevant to CALD premorbid-IQ estimation indicates that education and place of birth can contribute substantially to prediction, but population-appropriate calibration remains necessary. | Years of education may not capture educational quality, literacy, language of instruction, interrupted schooling, migration history, socioeconomic context, or differences in occupational classification. | (Barona et al., 1984; Crawford et al., 1989; Franzen et al., 1997; Sierra Sanjurjo et al., 2015) |
| Hold/Don't Hold Tests | Reliability depends on the underlying subtests; there is no single reliability coefficient for the approach. Estimates may also be affected when the presumed "hold" subtests are altered by current neurological disease. | In bilingual Mexican Americans, prediction by candidate hold measures differed across the U.S. and México WAIS-IV versions; in older Brazilian adults, Vocabulary showed relative stability across cognitively healthy, amnestic MCI, and Alzheimer's disease groups. | Crystallized knowledge and verbal subtests are influenced by educational opportunity, language, and cultural familiarity; disease-related change may reduce presumed hold performance. | (Bright & van der Linde, 2020; de Oliveira et al., 2014; Ontiveros & Gasquoine, 2024; Powell et al., 2003) |

**Table 1.** *Cont.*

| Method | Evidence on Reliability | Evidence on Validity | Key Population Limitations | Representative Evidence |
| --- | --- | --- | --- | --- |
| Word-Reading and Related Lexical Measures | Many language-specific reading measures report good internal consistency, interrater reliability, or test–retest stability, although stability can diminish with greater neurological disease severity. | This approach has the largest body of validation evidence across multiple populations and languages in the present review. Language-specific word-reading measures in Spanish, Portuguese, Japanese, Swedish, and Korean show associations with intellectual-ability criteria and/or relative stability in MCI or dementia; Brazilian Portuguese lexical-decision evidence provides a related language-based approach. Preservation becomes less dependable with greater dementia severity. | Highly dependent on literacy, educational quality and continuity, language proficiency, orthography, age of language acquisition, and sensory or speech functioning. English irregular-word paradigms cannot simply be translated across writing systems. | (Alves et al., 2012, 2013; Burin et al., 2000; Del Pino et al., 2018; Del Ser et al., 1997; Matsuoka et al., 2006; Olsen et al., 2015; Overman et al., 2021; Rolstad et al., 2008; Schrauf et al., 2006; Serrao et al., 2015; Sierra Sanjurjo et al., 2015; Yi et al., 2017) |
| Combination Methods | Reliability depends on the component measures and the stability of the regression model; there is no single reliability coefficient for the approach. | Combining demographic and performance-based predictors can improve prediction in some samples, including Latin American samples, but incremental validity is not uniform and combining predictors does not itself establish cultural validity. | Inherits limitations from each component measure; locally inappropriate demographic coding or reading measures can remain biased. Regression shrinkage can reduce accuracy at IQ extremes. | (Kirton et al., 2020; Powell et al., 2003; Shura et al., 2022; Sierra Sanjurjo et al., 2015; Watt et al., 2018) |
| Culturally Grounded Test Development | Reliability depends on the instrument and validation population. Demonstrating reliability for a culturally grounded cognitive measure does not establish reliability as a premorbid-IQ estimator. | KICA and CICA were developed for culturally appropriate cognitive assessment rather than premorbid-IQ estimation. Their relevance is methodological: they illustrate development of locally meaningful stimuli, task demands, and functional content. Related work on contextually important knowledge further illustrates that conventional academic measures may miss culturally valued acquired skills. | Highly context-specific; materials developed for one Indigenous or cultural community should not be generalized to another without local development and validation. | (Jacklin et al., 2020; LoGiudice et al., 2006; Sternberg et al., 2001) |

**Table 1.** *Cont.*

| Method | Evidence on Reliability | Evidence on Validity | Key Population Limitations | Representative Evidence |
|---|---|---|---|---|
| Informant-Based Approaches | The PCAS demonstrated high internal consistency (α = 0.90) and excellent inter-informant reliability (ICC = 0.96) in its Brazilian validation sample. | PCAS scores were associated with global cognitive functioning and reading ability and did not decrease across cognitively normal, mild cognitive impairment, and mild dementia groups, supporting its use as a structured estimate of premorbid cognitive abilities in that population. | Evidence is concentrated in a low-educated Brazilian older-adult sample; use depends on access to a knowledgeable informant; broader cultural, linguistic, and educational validation is needed. | (Apolinario et al., 2013) |
| Machine Learning and Normative Modeling | Conventional test reliability is not the primary property of these models. Reproducibility depends on model specification, data quality, internal validation, and successful external replication. | These methods can model complex or nonlinear relationships among demographic, linguistic, educational, and sociocultural variables, but the evidence reviewed is methodological or adjacent rather than direct validation of a clinical CALD premorbid-IQ estimator. | Training-sample underrepresentation; subgroup calibration and fairness; limited external validation; interpretability; and transportability across languages, settings, and populations. | (delCacho-Tena et al., 2024; Kiselica et al., 2024; Marquand et al., 2016; Rivera et al., 2024; Wolfers et al., 2018; Yuan et al., 2023) |
| Longitudinal and Dynamic Modeling | Conventional test–retest reliability is not the primary issue. Interpretability of repeated measurements depends on practice effects, selective attrition and survival, nonrandom missingness, stable measurement properties, and longitudinal measurement invariance. | Longitudinal models can characterize person-specific cognitive trajectories and dynamic prediction, but the studies discussed in this review illustrate the modeling approach rather than validate a premorbid-IQ estimator for CALD older adults. | Limited longitudinal CALD datasets; cohort effects; practice effects; attrition and survival bias; missing data; changes in test versions or translations; and measurement equivalence across languages and cultures. | (Bertola et al., 2021; Calamia et al., 2012; Chen et al., 2026; Gerstorf et al., 2011; Weuve et al., 2015; Widaman et al., 2010; Wu et al., 2022) |
| Harmonization | Reliability is not an independent property of harmonization itself. The quality of harmonized scores depends on the reliability of source measures and the adequacy of linking, calibration, or co-calibration procedures. | Harmonization can place scores from different instruments on common metrics and support pooled or cross-study analyses. It does not by itself establish cultural validity, measurement equivalence, or the validity of an individual premorbid-IQ estimate. | Differential item functioning; translation and linguistic equivalence; construct non-equivalence; and heterogeneity across source measures, cohorts, languages, and educational contexts must be evaluated explicitly. | (Briceño et al., 2023; Merkley et al., 2023; Mukherjee et al., 2023; Vonk et al., 2022) |

Note. Evidence summaries reflect reported test–retest, interrater, or other relevant reliability information; associations with independently measured IQ or other relevant criterion measures; stability in relevant neurological conditions; and external or cross-population validation where available. These descriptions constitute a narrative synthesis and should not be interpreted as formal evidence grades. Rows describing culturally grounded test development, machine learning/normative modeling, longitudinal/dynamic modeling, and harmonization include methodological or emerging research approaches; inclusion does not imply that each is a validated clinical premorbid-IQ estimator. CALD = culturally and linguistically diverse; CICA = Canadian Indigenous Cognitive Assessment; ICC = intraclass correlation coefficient; KICA = Kimberley Indigenous Cognitive Assessment; MCI = mild cognitive impairment; PCAS = Premorbid Cognitive Abilities Scale; WAIS = Wechsler Adult Intelligence Scale.

### 3.1. Clinical Judgment

Clinical judgment is one of the oldest approaches to estimating premorbid IQ and typically involves integrating information such as educational attainment, occupational history, demographic characteristics, observed behavior, and available cognitive test performance to formulate an estimate of an individual's prior level of functioning (Franzen et al., 1997). Its principal advantage is flexibility: clinicians can consider contextual information that may not be adequately represented in standardized prediction equations or performance-based measures. This may be particularly relevant for CALD older adults, for whom years of education, occupational status, literacy, language history, migration experiences, and other demographic characteristics may not have the same meaning or relationship to cognitive performance as they do in the populations on which conventional premorbid estimation methods were developed.

At the same time, unstructured clinical judgment raises concerns regarding reproducibility. In a direct comparison of clinicians' estimates with demographic regression equations in a sample of 60 healthy participants from the United Kingdom, Crawford et al. (2001) found that clinicians were more likely to overestimate IQ. Demographic variables (age, education, and social class) were more strongly correlated with obtained IQ ($r = 0.76$) than were average clinician estimates ($r = 0.64$). The observed directional error in that study was overestimation, and the broader concern is that clinicians may weight demographic and contextual information inconsistently across cases. Kareken (1997) similarly described limitations of unaided clinical inference and emphasized the advantages of actuarial information. This does not render clinical information irrelevant; rather, it supports using contextual information within an explicit framework instead of relying on unaided impression.

For CALD older adults, clinical judgment may therefore be most useful as a structured interpretive component of a multimethod approach rather than as a stand-alone method for generating a numerical premorbid IQ estimate. Clinicians can systematically document factors such as educational quality and interruptions, literacy opportunities, language acquisition and dominance, occupational complexity, migration history, and relevant sociocultural or sensory factors and use this information to evaluate the applicability of quantitatively derived estimates. However, in the absence of empirically validated weighting schemes for these contextual variables, clinicians should avoid assigning informal numerical corrections or treating particular cultural or linguistic characteristics as having fixed effects on estimated ability. When different sources of information yield discrepant estimates, the discrepancy and resulting uncertainty should be documented explicitly rather than resolved through unsupported numerical adjustment.

### 3.2. Best Performance Method

Although informal clinical judgment remains common in neuropsychological practice, several structured approaches have been developed to improve the consistency and reproducibility of premorbid IQ estimation. One of the earliest of these is the Best Performance Method (Mortensen et al., 1991), which attempts to infer premorbid IQ from an individual's highest observed level of cognitive performance. This method involves obtaining data from several sources, such as test scores, observations, interviews, informant (e.g., family) reports, and historical data. Using this approach, the data source that results in the highest observed estimate of ability is viewed as the most likely level of one's premorbid IQ. Relying on the assumption that high test scores likely reflect intact function, this approach to estimating premorbid IQ examines the scores on the various subtests of the IQ battery and uses the highest (i.e., best) performance as the baseline ability level; current functioning is then compared to this best performance and significant deviations are interpreted as declines from premorbid levels (Bright & van der Linde, 2020). Arguments against use of this

approach suggest significant overestimation of ability in light of the normal heterogeneity observed across subtest scores in healthy populations (Franzen et al., 1997).

### 3.3. Demographic-Only Regression Equations

Demographic-only regression models estimate premorbid IQ using demographic characteristics—typically age, education, sex, occupation, race/ethnicity, and/or geographic region—based on empirically derived relationships between these variables and IQ (e.g., Barona et al., 1984; Crawford et al., 1989; Wilson et al., 1978). An important advantage of these approaches is that the predictor variables are generally inexpensive and straightforward to obtain and, unlike performance-based measures, are not directly dependent on the patient's current cognitive performance. Demographic approaches may therefore provide a useful independent estimate when reading ability, language proficiency, speech or motor impairment, sensory limitations, or current neurological dysfunction compromises performance-based measures. They also provide an explicit and reproducible method for weighting demographic characteristics rather than requiring clinicians to determine informally how much weight to assign each factor.

One of the most commonly used demographic approaches is the Barona formula (Barona et al., 1984), a weighted regression method developed to estimate IQ on the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). The formula uses coded demographic characteristics, including age, sex, race, education, occupation, and geographic region, with each predictor assigned a different regression weight. Thus, the Barona approach converts information generally available through the clinical interview into a quantitative estimate of premorbid IQ. The Barona formula has been widely studied as a practical and inexpensive approach to estimating premorbid IQ. More recently, Kirton et al. (2020) examined its application to the WAIS-IV and found that, after correction for the Flynn effect, the Barona estimate showed the smallest mean difference from obtained WAIS-IV Full Scale IQ among the prediction models examined. This finding illustrates the potential utility of demographic estimation when differences in normative era are explicitly considered.

These advantages do not eliminate important limitations. The Barona Index has a relatively large standard error of estimation when applied at the individual level (Barona et al., 1984; Spreen & Strauss, 1998). Its application can also be complicated by decisions about how occupational and educational histories are classified and by the tendency of regression-based estimates to cluster toward the middle of the IQ distribution (Hawkins, 1995). In addition, because the equation was developed using an earlier U.S. normative sample, adjustments for normative changes over time, including the Flynn effect, may be necessary when it is applied to more contemporary measures (e.g., Kirton et al., 2020). More generally, regression to the mean can reduce accuracy at the extremes of the IQ distribution (see Reynolds, 1997; Veiel & Koopman, 2001). Specifically, prior literature has reported that the formula may overestimate premorbid IQ in individuals with actual IQs below 89 and underestimate premorbid IQ in individuals with actual IQs above 110 (Franzen et al., 1997).

The applicability of demographic equations also depends on how well the development sample and the operationalization of demographic predictors correspond to the individual being evaluated. This issue is particularly relevant for CALD populations. In a Spanish-speaking sample of 240 healthy adults residing in Colombia or Argentina, Sierra Sanjurjo et al. (2015) examined the utility of education, place of birth, occupation, and gender for predicting IQ as measured by the WAIS-III. Only education and place of birth contributed to the statistical models, jointly explaining 63.2% of the variance in Full Scale IQ. Although the demographics-only model accurately categorized 77.3% of individuals in the average IQ group, classification accuracy was considerably lower for the extremely low (34.9%), low-

average (17.1%), and high-IQ (12.5%) groups. These findings illustrate both the potential predictive value of demographic information and its population- and ability-level specificity.

Educational attainment presents an additional limitation because demographic formulas generally quantify the amount of education without adequately capturing its quality. Years of schooling are a crude proxy for educational experience. For many older adults, particularly those educated in under-resourced settings or whose educational experiences were affected by geographic location, migration, socioeconomic circumstances, or school disruption, years of education may not capture instructional quality, literacy opportunities, language of instruction, or interruptions in schooling. Accordingly, demographic regression equations can provide a useful independent anchor within a multimethod estimate of premorbid IQ, but their interpretation should consider the fit between the individual being evaluated and the population, historical context, and coding assumptions on which the equation was developed.

### *3.4. Hold/Don't Hold Tests*

Because demographic characteristics explain only part of the variance in IQ, investigators have also examined whether relatively preserved cognitive abilities can serve as direct indicators of premorbid IQ. This work led to the development of so-called "hold" tests. This approach is based on the notion that certain IQ subtests (typically from the WAIS) are resilient to effects from brain insult, allowing these tests that "hold" to be used for estimation of premorbid IQ (Lezak et al., 2004; Powell et al., 2003). The Vocabulary, Matrix Reasoning, Information, and Picture Completion subtests of the WAIS batteries are considered embedded "hold" tests (Bright & van der Linde, 2020). This approach can be applied in several ways, either by relying on a singular subtest (e.g., Vocabulary or the one with the highest score) or by averaging two or more of the "hold" tests (e.g., Vocabulary + Information/2). A limitation of this approach, though, is that it may be psychometrically flawed given that the premorbid IQ estimate is calculated from a subset of a battery used to calculate current IQ; this increases the likelihood of high predictive accuracy in neurologically healthy individuals while raising questions about its validity when used with neurological patients whose clinical conditions can affect performance on these measures. As with other approaches, this approach additionally fails to consider inter-subtest heterogeneity in healthy populations (Bright & van der Linde, 2020).

Direct evidence relevant to CALD premorbid-IQ estimation also suggests that the usefulness of candidate hold measures may depend on the language and normative framework of the intellectual measure being predicted. Ontiveros and Gasquoine (2024) examined five candidate hold measures—Information, Vocabulary, Matrix Reasoning, picture vocabulary, and the Test of Premorbid Functioning/Word Accentuation Test—in 60 neurologically intact, balanced bilingual Mexican American adults who completed both the U.S. WAIS-IV and WAIS-IV México in a counterbalanced repeated-measures design. All five hold measures significantly predicted summary scores on the U.S. WAIS-IV, whereas all except the Word Accentuation Test significantly predicted summary scores on the WAIS-IV México. Mean scores also differed systematically between the two WAIS-IV versions. These findings illustrate that measures assumed to reflect relatively preserved abilities cannot be treated as culturally or linguistically interchangeable and that their predictive validity may depend on the language, test version, educational background, and normative population to which they are applied.

Evidence from an older Brazilian sample with heterogeneous educational attainment also provides support for the relative stability of Vocabulary as a hold measure. de Oliveira et al. (2014) found that WAIS-III Vocabulary performance did not differ significantly among cognitively healthy older adults and individuals with amnestic mild cognitive impairment

or Alzheimer's disease, whereas measures of fluid intellectual ability showed greater decline across diagnostic groups. These findings support the potential utility of Vocabulary as an indicator of prior ability in this population, although they do not establish its validity as a culturally invariant or stand-alone estimate of Full Scale IQ.

### 3.5. Word Reading and Related Lexical Measures

Among standalone performance-based approaches, oral word-reading measures are among the most widely studied methods for estimating premorbid IQ. Common examples include the National Adult Reading Test (NART; Nelson, 1982), the Wide Range Achievement Test–Fifth Edition (WRAT-5) Word Reading subtest (Wilkinson & Robertson, 2017), and the Test of Premorbid Functioning (TOPF; NCS Pearson, 2009). These measures are brief, inexpensive, and based on highly practiced reading skills that are generally more closely associated with crystallized than fluid intellectual abilities. In a comparison of premorbid estimation approaches, Bright and van der Linde (2020) found that NART performance correlated with WAIS-IV Full Scale IQ ($r = 0.69$), with the association largely reflected in Verbal Comprehension ($r = 0.66$). Word-reading estimates were also more precise than estimates derived from the Best Performance Method.

Evidence for reading-based estimation extends beyond English-language measures. In 74 older adults from Buenos Aires, Burin et al. (2000) found support for the concurrent validity of the Spanish Word Accentuation Test (WAT) based on its association with WAIS Vocabulary performance and years of education. In 80 older Spanish-dominant immigrants living in the United States, Schrauf et al. (2006) found that the WAT accounted for substantial variance in performance on the Batería Woodcock-Muñoz Revisada and Raven's Colored Progressive Matrices. Similarly, Krueger et al. (2006) found that WAT performance accounted for an additional 5% of variance in WAIS-III performance after controlling for age and education. In adults from Argentina and Colombia, a revised WAT correlated strongly with WAIS-III IQ ($r = 0.82$), and a combined model incorporating WAT-R performance, education, and place of birth explained 76% of IQ variance (Sierra Sanjurjo et al., 2015). Del Pino et al. (2018) subsequently developed demographically calibrated norms for the WAT and a Pseudo-Words Reading subtest in 700 healthy adults from Spain aged 18–86. Age and education significantly influenced performance across the premorbid measures, whereas sex did not; together, age and education accounted for 1.9–33.2% of variance across WAT and pseudo-word outcomes. The Pseudo-Words Reading subtest also appeared more resistant to age-related decline than the WAT. These findings further underscore that even language-specific premorbid estimators require population-appropriate demographic calibration rather than being interpreted independently of educational and demographic context (Del Pino et al., 2018).

Comparable evidence has emerged in other languages. Among Portuguese speakers, the TeLPI was validated against WAIS-III Full Scale IQ, with the resulting prediction model accounting for 63% of FSIQ variance (Alves et al., 2012). Subsequent validation showed that TeLPI performance remained relatively stable across cognitively healthy, mild cognitive impairment (MCI), and Alzheimer's disease groups after accounting for education (Alves et al., 2013). Brazilian Portuguese evidence also includes a lexical-decision approach. Serrao et al. (2015) evaluated a Lexical Decision Test (LDT) in 38 cognitively healthy older adults, 61 individuals with mild cognitive impairment, and 26 individuals with Alzheimer's disease. LDT performance did not differ significantly across the three groups, despite group differences in several measures of current cognitive functioning and contemporaneously estimated IQ. In the cognitively healthy group, LDT performance together with years of schooling explained 66.3% of the variance in estimated IQ ($R^2 = 0.663$). These findings provide additional support for a language-specific lexical approach to premorbid estima-

tion in Brazilian Portuguese, although the criterion was contemporaneously estimated IQ rather than independently observed premorbid ability and further validation is needed. In Korean-speaking older adults, the Korean Adult Reading Test (KART) predicted Full Scale IQ with $R^2 = 0.63$ in its validation sample, and KART-predicted IQ did not significantly differ across cognitively normal, MCI, and Alzheimer's disease groups despite marked differences in current cognition (Yi et al., 2017).

The linguistic limitations of reading-based estimators extend beyond differences between opaque and relatively transparent alphabetic orthographies. An estimator based on irregular English word pronunciation cannot simply be translated into a writing system with different orthographic, morphological, and phonological relationships. This is relevant to languages using morphosyllabic characters, such as Chinese, as well as languages using abjad scripts, such as Arabic. The existence of successful language-specific measures, including the Japanese Adult Reading Test (JART; Matsuoka et al., 2006), Swedish National Adult Reading Test (NART-SWE; Rolstad et al., 2008), Korean Adult Reading Test (KART; Yi et al., 2017), and Spanish and Portuguese adaptations of irregular-word paradigms, suggests that reading-based estimation is not inherently restricted to English. Rather, the task and items must be developed around the orthographic, phonological, and lexical characteristics of the target language and validated in the population in which the measure will be used. Notably, the Japanese and Swedish measures also showed relative preservation in older adults with Alzheimer's disease, although such findings remain specific to the populations, languages, and disease stages studied.

Longitudinal evidence also provides support for the relative stability of word-reading ability. In a predominantly racial/ethnic minority sample of adults with HIV, WRAT-3 Reading Recognition demonstrated high test-retest reliability over an average interval of nearly five years (ICC = 0.97), with coefficients exceeding 0.92 across demographic, medical, and neurocognitive subgroups (Olsen et al., 2015). Such findings support the use of reading as a relatively stable indicator of prior ability in some clinical populations, but they should not be interpreted to mean that word reading is universally preserved.

This qualification is particularly important in neurodegenerative disease. A systematic review by Overman et al. (2021) found cautious support for word-reading measures in mild dementia across several languages, but substantially weaker evidence in more advanced disease. More recently, Marier et al. (2024) found that irregular-word reading differed significantly across the Alzheimer's disease continuum, including early and late mild cognitive impairment and Alzheimer's disease (partial $\varepsilon^2 = 0.09$), with poorer performance associated with greater cognitive and semantic decline. These findings suggest that word-reading measures may underestimate premorbid IQ when the disease process affects the semantic systems required for irregular-word pronunciation.

Educational experience further complicates interpretation. Sayegh et al. (2014) found that self-rated high school educational quality was moderately associated with WRAT-4 Word Reading ($r = 0.37$). After accounting for race/ethnicity, socioeconomic status, and years of education, self-rated high school educational quality explained an additional 3% of variance in reading performance ($\Delta R^2 = 0.03$; standardized $\beta = 0.18$). Thus, reading performance reflects meaningful aspects of educational experience, but reading-based estimates should not be assumed to be unaffected by ethnocultural, linguistic, or normative context (Cosentino et al., 2007; Sayegh et al., 2014).

More broadly, educational quality may have implications for later-life cognitive functioning beyond its association with reading performance. Using Health and Retirement Study data linked to historical state-level indicators of educational quality, Ailshire et al. (2025) found that older adults who had been educated in better-resourced state education systems had lower prevalence of dementia and cognitive impairment without dementia.

Educational attainment alone did not account for the higher prevalence of cognitive impairment observed in the U.S. South, whereas differences in state-level educational quality accounted for these regional disparities. Although these findings do not directly validate a premorbid-IQ estimator, they further illustrate why years of schooling alone may inadequately characterize the cognitive implications of educational experience.

Sensory, speech, and motor changes may further limit the validity of word-reading estimates in older adults. Visual impairment can interfere with accurate perception of written stimuli, while speech or motor changes may affect oral production independently of premorbid IQ. Sensory limitations may make testing more effortful and may require appropriate accommodations (Abdelsalam et al., 2025; Dumassais et al., 2024). Separately, older adults may be vulnerable to age-based stereotype threat, which can adversely affect cognitive performance through anxiety and interfering thoughts related to negative stereotypes about aging and cognitive decline (Régner & Huguet, 2025). Appropriate sensory, speech, or motor accommodations should therefore be considered when needed, with departures from standardized administration documented and incorporated into interpretation (Dumassais et al., 2024; Gosselt et al., 2020; Hill-Briggs et al., 2007; Magasi et al., 2017).

Taken together, word-reading measures can provide useful estimates of premorbid IQ, particularly when supported by appropriate linguistic and within-population validation. However, their accuracy depends on the integrity of the language, reading, sensory, and motor systems required for performance. In CALD older adults, word-reading estimates should therefore be interpreted in light of disease severity, language proficiency, literacy, educational quality and continuity, sensory and motor functioning, and the linguistic and orthographic properties of the language for which the measure was developed and validated.

### *3.6. Combination Methods*

Because demographic and performance-based approaches capture partly distinct sources of information, combination methods have been developed to integrate these predictors within a single estimate. These approaches most commonly combine word-reading performance with demographic characteristics, reflecting evidence that each accounts for different sources of variance in IQ. However, the incremental benefit of combining methods is not uniform. Bright and van der Linde (2020) found that word-reading measures provided relatively strong estimates of WAIS-IV Full Scale IQ, with the addition of demographic variables yielding only modest improvement, whereas Kirton et al. (2020) found that the relative performance of TOPF, OPIE-3, and demographic approaches varied depending on the criterion used to evaluate accuracy. Thus, combination methods may improve estimation in some contexts, but no single approach can be assumed to provide the most accurate estimate across individuals, ability levels, or clinical populations.

Watt et al. (2018) developed regression equations for predicting premorbid functioning in an Australian sample using NART performance alone and in combination with demographic variables. NART performance independently accounted for a substantial proportion of variance in WAIS-IV Full Scale IQ, while the addition of age, sex, and education produced a modest but meaningful improvement in predictive accuracy. Importantly, when the equations were applied to a neurologically impaired sample, models based solely on NART performance significantly overestimated premorbid IQ, whereas combined models incorporating demographic variables did not. Similarly, demographic-corrected estimates derived from the TOPF demonstrated greater accuracy than uncorrected TOPF estimates. These findings suggest that demographic variables may provide incremental validity beyond oral word-reading performance, particularly when locally derived regression equations are available. However, because these equations were developed and validated in predomi-

nantly English-speaking Australian samples, their generalizability to CALD populations remains uncertain.

Nevertheless, findings from other combination methods have been less consistent. In a sample of individuals with known neurological diagnoses, Powell et al. (2003) found the Barona formula to be more sensitive to cognitive decline than the Oklahoma Premorbid Intelligence Estimate (OPIE; Krull et al., 1995). The OPIE combines demographic variables (age, education, race, and occupation) with current performance on two presumed "hold" measures, WAIS-R Vocabulary and Picture Completion. Powell et al. suggested that the relative vulnerability of these performance measures to neurological impairment may cause the OPIE to reflect current cognitive functioning rather than premorbid ability in some clinical populations. The OPIE-3 subsequently extended this combined approach to the WAIS-III using demographic variables together with various combinations of Vocabulary, Information, Matrix Reasoning, and Picture Completion performance (Schoenberg et al., 2002). In a later comparison with WAIS-IV Full Scale IQ, Kirton et al. (2020) found that the OPIE-3 correlated with obtained IQ but tended to overestimate it, whereas the Flynn-adjusted Barona equation produced the smallest mean difference from obtained Full Scale IQ.

Studies combining demographic variables with oral word reading tests have likewise yielded conflicting results. Proprietary equations available through Pearson's Advanced Clinical Solutions WAIS-IV and WMS-IV (ACS; NCS Pearson, 2009) software allow clinicians to estimate premorbid functioning using various combinations of TOPF performance and demographic characteristics. However, in a neurologically healthy veteran sample, Shura et al. (2022) found that the three TOPF scores incorporating word-reading performance differed significantly from obtained WAIS-IV Full Scale IQ, whereas the two scores based solely on demographic information did not. Across all five TOPF scores, discrepancies were greatest at the extremes of the IQ distribution, although the authors noted the small subsamples at those extremes. Conversely, Bright and van der Linde (2020) reported that oral word reading tests provided the most reliable and precise estimates of Full Scale IQ in neurologically healthy adults, with demographic variables contributing only modest improvements in predictive accuracy.

Comparable findings have also been reported in non-English-speaking populations. In a Spanish-speaking sample, Sierra Sanjurjo et al. (2015) found that a revised WAT alone accurately classified individuals with average IQ (90.9%) but showed substantially lower accuracy among individuals with extremely low (65.1%), low-average (60.4%), and high IQ (0%) scores. A combined model incorporating both oral word reading performance and demographic variables produced the highest overall classification accuracy; however, meaningful limitations remained (extremely low IQ = 72.1%, low average IQ = 60.4%, average IQ = 86.4%, and high IQ = 12.5%).

Overall, the available literature suggests that combining demographic information with performance-based measures may produce modest improvements in premorbid IQ estimation under certain circumstances, particularly when regression equations are developed and validated within the population to which they are applied. However, the incremental gains are generally small, prediction errors remain greatest at the extremes of the IQ distribution, and the validity of existing combination methods outside the populations in which they were derived—including CALD older adults—remains largely unestablished.

### *3.7. Culturally Grounded and Contextual Approaches*

In addition to conventional psychometric approaches, several lines of work illustrate ways of incorporating culturally relevant content and systematically collected contextual information into the assessment of prior cognitive functioning. These approaches differ

in important respects: some provide methodological models for developing culturally grounded assessment tools, whereas others use structured informant information to characterize premorbid abilities. Although these methods are not interchangeable and vary in the extent to which they directly estimate premorbid IQ, they are relevant to the development of more culturally responsive approaches for CALD older adults.

#### 3.7.1. Culturally Grounded Test Development

Recent developments in neuropsychological assessment have increasingly emphasized the limitations of applying conventional premorbid IQ estimation approaches across CALD populations. Traditional normative methods, including demographic corrections and reading-based estimators, were primarily developed in Western, English-speaking, and highly educated cohorts, raising concerns regarding their transportability to multilingual, migrant, and socioeconomically heterogeneous older adults (Ardila, 2005; Manly, 2008; Nielsen, 2022). Cross-cultural neuropsychological research and professional guidance further emphasize that interpretation of cognitive test performance may be affected by educational quality and literacy, acculturation and language proficiency, migration history, and broader sociocultural context (Manly et al., 2002; Mendoza et al., 2022; Nielsen et al., 2024; Rivera Mindt et al., 2010). Consequently, the application of norms derived from majority populations may contribute to systematic misclassification and diagnostic inequities among older CALD individuals whose educational and linguistic experiences differ substantially from normative reference samples (Brickman et al., 2006; Gasquoine, 2009; Merkley et al., 2023).

More recent cultural neuropsychology models increasingly frame cognition as embedded within sociocultural and historical experience, rather than viewing culture primarily as a demographic correction factor (Judd et al., 2009). Within this framework, culturally grounded assessment extends beyond simple translation or demographic correction and instead emphasizes the interaction among lived experience, language ecology, educational opportunity, migration history, and neurocognitive functioning (Rivera Mindt et al., 2010). However, implementation challenges remain substantial. Many culturally adjusted normative systems continue to rely on broad racial, ethnic, or language categories that inadequately capture within-group heterogeneity, particularly among aging immigrant populations (Gasquoine, 2009; Manly & Echemendia, 2007). Furthermore, ongoing debates remain regarding whether demographically adjusted or race/ethnicity-specific norms improve diagnostic fairness or instead risk normalizing cognitive disparities that partly reflect structural inequities and unequal educational opportunities (Byrd & Rivera Mindt, 2022; Manly & Echemendia, 2007; Possin et al., 2021).

Importantly, administering an instrument in an examinee's preferred language does not by itself establish the appropriateness of its normative reference group. Ontiveros and Gasquoine (2024), for example, directly compared the U.S. and México versions of the WAIS-IV in balanced bilingual Mexican American adults and found systematic differences in scores across versions. Educational history also appeared relevant to the pattern of performance. These findings illustrate that language, national origin, educational experience, and the characteristics of the normative sample may interact in ways that complicate the selection and interpretation of intelligence measures in CALD examinees. Accordingly, a translated or nationally adapted version of a measure should not automatically be assumed to provide the most appropriate normative comparison for a bilingual or immigrant examinee.

Community-engaged development of cognitive measures for Indigenous populations provides an important example of culturally grounded test construction. Although these instruments were not designed to estimate premorbid IQ, they illustrate how assessment content, task demands, language, and functional relevance can be developed around the

lived cultural environment of the population being assessed. Standardized mainstream cognitive screening instruments may be culturally, linguistically, and educationally inappropriate for some Indigenous populations. In response, community-engaged approaches have been used to develop culturally grounded assessment tools. The Kimberley Indigenous Cognitive Assessment (KICA; LoGiudice et al., 2006), for example, was developed for older Aboriginal Australians in remote communities, while the Canadian Indigenous Cognitive Assessment (CICA; Jacklin et al., 2020) was created collaboratively with First Nations communities on Manitoulin Island and incorporates locally relevant imagery, activities, and language (Jacklin et al., 2020; LoGiudice et al., 2006). Both measures have demonstrated utility as culturally adapted cognitive assessment tools, although neither was designed to estimate premorbid IQ.

Measures such as the KICA (LoGiudice et al., 2006) and CICA (Jacklin et al., 2020) are not premorbid-IQ estimators and should not be interpreted as such. Their relevance to the present review is methodological. These instruments illustrate an approach to cognitive assessment in which stimuli, task demands, communication practices, and functional content are developed around the lived cultural environment of the population being assessed (Jacklin et al., 2020; LoGiudice et al., 2006). For future premorbid-IQ measures, this principle suggests that crystallized knowledge may need to be operationalized using culturally familiar knowledge and skills rather than assuming that the same vocabulary, reading, educational, or occupational indicators are equivalent across populations. Indigenous assessment measures therefore provide models of culturally grounded test development rather than substitutes for validated measures of premorbid IQ.

Sternberg et al. (2001) provide another useful illustration of the distinction between practical, contextually relevant knowledge and conventional academic intelligence when considering the limitations of traditional approaches to estimating premorbid IQ in CALD populations. Their research demonstrated that children may develop highly adaptive, contextually relevant knowledge and problem-solving abilities that are measurable within their cultural environments but are not necessarily reflected in conventional academic intelligence measures. Specifically, children in rural Kenya demonstrated substantial tacit knowledge concerning natural herbal medicines that was viewed within their communities as important for adaptation to their environment, yet this practical knowledge was unrelated—and in some cases inversely related—to conventional measures of academic intelligence and achievement. These findings suggest that conventional measures of IQ may capture skills developed through formal educational experiences while overlooking other forms of acquired knowledge and adaptive expertise. Although Sternberg and colleagues' work was conducted in a pediatric population, the distinction between contextually important and academically valued skills may have important implications for neuropsychological assessment of CALD older adults. Future research could examine whether culturally relevant indicators of acquired knowledge, occupational expertise, adaptive functioning, or language-specific skills may provide additional sources of information for estimating premorbid IQ beyond traditional reading-based and educational attainment approaches.

#### 3.7.2. Informant-Based Approaches

A complementary approach is illustrated by the Premorbid Cognitive Abilities Scale (PCAS; Apolinario et al., 2013), an informant-based measure developed to estimate premorbid cognitive abilities in older adults with low levels of education. Rather than relying on an unstructured clinician impression, the PCAS systematically assesses educational attainment, major lifetime occupation, reading and writing abilities, reading habits, calculation abilities, use of commonly available technology, and the ability to locate specific information. In a Brazilian validation sample of 132 older adults comprising cognitively

normal participants, individuals with mild cognitive impairment, and individuals with mild dementia, PCAS scores were strongly associated with global cognitive functioning ($r = 0.73$) and reading ability ($r = 0.82$). The scale also demonstrated high internal consistency (Cronbach's $\alpha = 0.90$) and excellent inter-informant reliability (ICC = 0.96), and scores did not decrease across progressively impaired diagnostic groups (Apolinario et al., 2013). Although validation in additional cultural, linguistic, and educational populations is needed, the PCAS is conceptually important because it demonstrates that contextual information about premorbid functioning can be collected systematically and reproducibly rather than incorporated solely through unstructured clinical judgment.

### 3.8. Machine Learning and Normative Modeling

Recent advances in normative modeling and machine learning may provide additional opportunities to improve individualized premorbid IQ estimation. Unlike traditional demographic regression equations, contemporary normative approaches can model complex and potentially nonlinear relationships between cognitive performance and variables such as age, education, language, and other individual characteristics (Marquand et al., 2016; Wolfers et al., 2018). Related work in neuropsychology has increasingly emphasized individualized and data-driven approaches to normative interpretation rather than relying exclusively on fixed group-based comparisons (delCacho-Tena et al., 2024; Kiselica et al., 2024; Rivera et al., 2024). In principle, models developed specifically for premorbid estimation could incorporate a broader set of variables relevant to CALD populations, including educational quality and continuity, literacy, language acquisition and dominance, migration history, occupational experience, and other sociocultural characteristics. More directly relevant to premorbid ability, Abdullah et al. (2026) used data from 2691 participants in the multiethnic Health and Aging Brain Study–Health Disparities and applied an XGBoost model incorporating social determinants of health to distinguish low versus average premorbid-ability groups defined by word-reading performance. The model achieved an area under the curve of 0.72, with worry, neighborhood deprivation, income, high school completion, and tangible social support among the most influential predictors. Because the target groups were themselves defined using word-reading scores rather than independently observed premorbid IQ, these findings illustrate the potential for machine-learning models to incorporate broader contextual variables but do not establish an independent machine-learning estimator of premorbid IQ.

Machine-learning approaches may be particularly useful when relationships among these variables are nonlinear or interactive. However, greater predictive complexity does not necessarily confer greater validity across populations. Model performance depends substantially on the characteristics of the samples used for training and validation, and aggregate predictive accuracy can obscure clinically important differences in performance across demographic groups. For example, Yuan et al. (2023) found that longitudinal machine-learning models predicting progression to mild cognitive impairment and Alzheimer's disease demonstrated notable disparities in sensitivity across racial and ethnic groups despite reasonable overall model performance. Such findings are particularly relevant to premorbid estimation in CALD populations, where underrepresentation in training datasets could reproduce rather than reduce existing measurement inequities.

Accordingly, future machine-learning approaches to premorbid IQ estimation should be externally validated in linguistically, culturally, educationally, and geographically heterogeneous samples and should report subgroup-specific discrimination, calibration, and prediction error rather than overall accuracy alone. Interpretability is also important because clinicians must be able to understand why a model generates a particular estimate and whether the variables driving that estimate are appropriate for the individual being

assessed. At present, machine-learning and normative-modeling approaches should therefore be regarded as promising research complements to established premorbid-estimation methods rather than validated clinical substitutes.

### 3.9. Longitudinal and Dynamic Modeling

A related but distinct direction involves longitudinal and dynamic approaches to cognitive prediction. Most conventional methods estimate premorbid IQ retrospectively from information obtained at a single assessment and therefore treat premorbid ability largely as a static quantity. Longitudinal methods instead permit cognitive functioning to be represented as an individual trajectory, potentially allowing models to distinguish stable between-person differences from within-person change and to characterize heterogeneity in rates and patterns of cognitive decline.

Recent studies provide concrete examples of this shift. Wu et al. (2022) used 16 years of longitudinal data from 3502 Chinese adults aged 65 to 105 years to identify heterogeneous cognitive trajectories and subsequently applied machine-learning methods to predict membership in unfavorable trajectory groups. More recently, Chen et al. (2026) developed a transformer-based time-series model using longitudinal China Health and Retirement Longitudinal Study (CHARLS) data to predict mild cognitive impairment risk among older adults with depressive symptoms and externally evaluated the model using the Chinese Longitudinal Healthy Longevity Survey (CLHLS). Their approach used repeated observations within a sliding-window framework so that temporal relationships among changing predictors contributed to prediction rather than treating each assessment occasion as independent. Neither study was designed to estimate premorbid IQ and therefore neither provides evidence for a validated premorbid estimator. They nevertheless illustrate how repeated, heterogeneous information can be incorporated into dynamic models of cognitive change.

Applied to premorbid estimation, related approaches could eventually shift the clinical question from estimating a single historical IQ score toward estimating the cognitive level and trajectory expected for an individual with a particular developmental, educational, linguistic, occupational, and sociocultural history and determining whether subsequent performance deviates meaningfully from that trajectory. Such models could be especially valuable when a single cross-sectional estimate is difficult to interpret because conventional assumptions regarding education, literacy, or language are not met.

Substantial methodological challenges remain before this approach could be used clinically. Longitudinal cognitive data are affected by practice effects, selective attrition and survival, nonrandom missingness, and cohort effects that may complicate the separation of true cognitive change from changes associated with repeated testing or historical and sociocultural context (Calamia et al., 2012; Gerstorf et al., 2011; Weuve et al., 2015). Changes in test versions, translations, or other measurement properties across study waves create additional difficulties. Longitudinal measurement invariance is therefore important for determining whether observed score changes represent change in the underlying cognitive construct rather than change in how that construct is measured (Bertola et al., 2021; Widaman et al., 2010). For CALD populations, future dynamic models should additionally evaluate measurement equivalence and predictive performance across languages, educational strata, cultural groups, and geographic settings. Dynamic modeling may ultimately provide a useful framework for individualized premorbid estimation, but it remains an emerging research direction rather than an established clinical method.

### 3.10. Harmonization

For purposes of this review, harmonization refers to procedural, qualitative, and statistical approaches used to make cognitive data collected across different tests, languages, cohorts, or cultural contexts sufficiently comparable for pooled or cross-study analysis while preserving the construct of interest and evaluating potential measurement non-equivalence (Briceño et al., 2023; Merkley et al., 2023; Mukherjee et al., 2023; Vonk et al., 2022). Approaches may include the prospective use of common or linking items, item-level calibration, and statistical co-calibration or latent-variable methods that place scores from different instruments on a common metric (Mukherjee et al., 2023; Vonk et al., 2022). In CALD populations, these procedures also require attention to linguistic and cultural equivalence because apparently similar or translated items may function differently across populations even when they are intended to measure the same cognitive construct (Briceño et al., 2023; Merkley et al., 2023; Vonk et al., 2022). Harmonization is therefore best understood as methodological infrastructure that may facilitate the development and validation of premorbid-IQ models across populations; it is not itself a method for estimating an individual's premorbid IQ.

This distinction is particularly relevant to cross-cultural neuropsychological research. Harmonization studies involving older adults have demonstrated that items judged conceptually similar across languages or populations may nevertheless show evidence of differential item functioning or other forms of measurement non-equivalence (Briceño et al., 2023; Vonk et al., 2022). At the same time, successful co-calibration across different cognitive batteries can place scores on common metrics and thereby permit larger pooled analyses and more direct comparisons across cohorts (Mukherjee et al., 2023). These methods could ultimately support development and validation of premorbid-IQ models in more heterogeneous populations by allowing evidence from linguistically, culturally, and educationally diverse cohorts to be integrated. However, harmonization cannot by itself establish that an estimator is culturally valid or clinically appropriate for a particular individual.

## 4. Comparative Evidence for Premorbid IQ Estimation in CALD Older Adults

The approaches reviewed in Section 3 vary considerably in the type and strength of evidence available regarding culturally, linguistically, educationally, and otherwise heterogeneous populations. To avoid conflating validation of an estimator within a particular population with evidence that diversity-related characteristics themselves affect estimation, we distinguish three forms of evidence. First, direct evidence on diversity-related influences includes studies that explicitly evaluate whether cultural, linguistic, educational, demographic, migration-related, or related characteristics affect the performance, calibration, stability, or predictive validity of a premorbid-IQ or premorbid-cognitive-ability estimator, or that directly compare estimator performance across relevant populations, languages, or normative contexts. Second, within-population validation evidence establishes the reliability, validity, or stability of an estimator within a particular language or population but does not necessarily test the influence of diversity-related characteristics. Third, contextual evidence addresses factors such as literacy, educational quality, acculturation, socioeconomic context, or cross-cultural test performance without directly validating a premorbid-IQ estimator. Section 4.1 focuses on direct evidence regarding diversity-related influences on premorbid-IQ estimation, Section 4.2 reviews empirical contextual evidence, and Section 4.3 synthesizes the implications across approaches. Within-population validation evidence remains important to the method-specific review in Section 3 and is distinguished accordingly in Table 2.

**Table 2.** Selected Empirical Studies Relevant to Premorbid-IQ Estimation Across Culturally, Linguistically, Educationally, or Migration-Defined Populations Discussed in the Present Narrative Review.

| Study | Population/ Context | Premorbid Estimation Approach | Main Finding | Relevance/Limitation |
| --- | --- | --- | --- | --- |
| Sierra Sanjurjo et al. (2015) | 240 healthy Spanish-speaking adults from Argentina and Colombia. | Compared demographic predictors, WAT-Revised, and a combined demographic + WAT-R model. | Education and place of birth explained 63.2% of Full Scale IQ variance; a combined model incorporating WAT-R, education, and place of birth explained 76%. Prediction/classification was weaker at IQ extremes. | Direct evidence on diversity-related influences: Demonstrates that education and place of birth contribute to IQ prediction and that a combined demographic and WAT-R model improves prediction, although accuracy remains weaker at IQ extremes. |
| Del Pino et al. (2018) | 700 healthy adults from Spain, ages 18–86; sample calibrated to age, education, and sex. | Demographically calibrated norms for the Word Accentuation Test (WAT) and Pseudo-Words Reading (PW) subtest. | Age and education significantly affected premorbid-measure performance, whereas sex did not. Across WAT and PW outcomes, age and education accounted for 1.9–33.2% of score variance; PW performance was more resistant to age-related decline than WAT performance. | Direct evidence on diversity-related influences: Demonstrates effects of age and education, but not sex, on Spanish-language premorbid measures and supports demographic calibration of language-specific estimators. |
| Ontiveros and Gasquoine (2024) | 60 neurologically intact, balanced bilingual Mexican American adults assessed with U.S. and México WAIS-IV versions in counterbalanced order. | Five candidate hold measures: WAIS-IV Information, Vocabulary, and Matrix Reasoning; picture vocabulary; and TOPF (English)/WAT (Spanish). | All hold measures significantly predicted U.S. WAIS-IV summary scores; all except the WAT predicted México-version summary scores. Mean México-version scores were generally higher than U.S.-version scores, with educational history affecting the pattern of score differences. | Direct evidence on diversity-related influences: Demonstrates that test version, language/national norms, and educational history can affect scores and predictive relationships in balanced bilingual Mexican American adults. Neurologically intact sample; not a dementia-specific validation study. |
| Olsen et al. (2015) | 88 English-speaking adults with advanced HIV; predominantly African American and Hispanic/Latino; mean age 46.9 years. | WRAT-3 Reading Recognition as a word-reading estimate of premorbid ability. | WRAT-3 Reading Recognition showed high long-term stability over an average retest interval of 4.9 years (ICC = 0.97), with coefficients above 0.92 across demographic, medical, and neurocognitive subgroups. | Direct evidence on diversity-related influences: Provides a useful null finding, with high long-term reading stability and no meaningful reliability differences across the demographic, medical, and neurocognitive subgroups examined. The sample was middle-aged rather than geriatric and disease-specific. |

**Table 2.** *Cont.*

| Study | Population/ Context | Premorbid Estimation Approach | Main Finding | Relevance/Limitation |
| --- | --- | --- | --- | --- |
| Schrauf et al. (2006) | Older Spanish-dominant immigrants living in the United States. | Expanded/adapted WAT for older U.S. Spanish-speaking immigrants. | WAT performance accounted for substantial variance in Batería Woodcock-Muñoz Revisada scores and was associated with Raven's Colored Progressive Matrices. | Within-population validation evidence: Evaluates the validity and potential need for adaptation of the Spanish WAT in older Spanish-dominant U.S. immigrants, supporting applicability of the original word list in this sample. The study did not directly test whether individual differences in acculturation, dialect, migration history, or other diversity-related characteristics affected estimator performance. |
| Del Ser et al. (1997) | Spanish-speaking adults in Spain; included application to dementia. | Word Accentuation Test (WAT), a Spanish reading-based estimator using low-frequency words presented without accent marks. | WAT performance showed strong association with an intellectual criterion derived from WAIS subtests ($r = 0.837$) and a moderate-to-strong association with Raven's Progressive Matrices ($r = 0.655$). | Within-population validation evidence: Foundational Spanish-language validation showing that reading-based premorbid estimation can be adapted to Spanish. Applicability remains specific to the population and language studied. |
| Burin et al. (2000) | 74 older adults from Buenos Aires, Argentina. | Buenos Aires version of the WAT. | Supported concurrent validity with verbal intellectual ability and education and showed high internal consistency; performance was relatively independent of age, memory, and executive functioning. | Within-population validation evidence: Extends WAT validation to an Argentine older-adult sample. Findings should not be generalized automatically to other Spanish-speaking populations. |
| Krueger et al. (2006) | Spanish-speaking adults in the United States. | Word Accentuation Test-Chicago/revised WAT approach. | WAT performance contributed an additional 5% of variance in intellectual performance after controlling for age and education. | Within-population validation evidence: Demonstrates incremental predictive value of WAT performance beyond age and education in U.S. Spanish-speaking adults. The study does not directly establish cross-population transportability. |
| Matsuoka et al. (2006) | Japanese older adults: 100 cognitively normal participants and 70 age-, sex-, and education-matched adults with Alzheimer's disease. | Japanese Adult Reading Test (JART), using irregular Kanji compound-word pronunciation to estimate premorbid IQ. | In the validation sample, JART-predicted Full Scale IQ fit observed IQ well ($R^2 = 0.78$). JART-predicted IQ did not differ significantly between cognitively normal and Alzheimer's disease groups despite lower observed IQ in the disease group. | Within-population validation evidence: Supports validity and relative disease stability of a Japanese reading-based estimator. Findings remain specific to Japanese speakers and the disease stages studied. |

**Table 2.** *Cont.*

| Study | Population/ Context | Premorbid Estimation Approach | Main Finding | Relevance/Limitation |
| --- | --- | --- | --- | --- |
| Rolstad et al. (2008) | Swedish-speaking healthy adults and patients with mild Alzheimer's disease. | Swedish National Adult Reading Test (NART-SWE), constructed with loan words because Swedish pronunciation rules are comparatively regular. | NART-SWE showed very high interrater and retest reliability and internal consistency, correlated strongly with IQ, and did not differ significantly between healthy controls and patients with mild Alzheimer's disease. | Within-population validation evidence: Provides Swedish-language psychometric and disease-stability evidence. Applicability beyond Swedish speakers and mild Alzheimer's disease requires separate validation. |
| Alves et al. (2012) | Portuguese-speaking adults. | Irregular Word Reading Test (TeLPI). | TeLPI was validated against WAIS-III Full Scale IQ; the prediction model accounted for 63% of FSIQ variance. | Within-population validation evidence: Supports validation of a Portuguese-language irregular-word reading estimator against WAIS-III Full Scale IQ. |
| Alves et al. (2013) | Portuguese-speaking cognitively healthy adults and individuals with MCI or Alzheimer's disease. | TeLPI. | After accounting for education, TeLPI performance remained relatively stable across cognitively healthy, MCI, and Alzheimer's disease groups. | Within-population validation evidence: Provides disease-stability evidence for TeLPI in Portuguese-speaking participants after accounting for education; generalization beyond the validation population requires caution. |
| Serrao et al. (2015) | 125 Brazilian older adults: 38 cognitively healthy, 61 with MCI, and 26 with Alzheimer's disease. | Brazilian Portuguese Lexical Decision Test (LDT), developed using regular words of varying frequency as a lexical-decision measure of premorbid IQ. | LDT performance did not differ significantly across cognitively healthy, MCI, and Alzheimer's disease groups, whereas Matrix Reasoning and contemporaneously estimated IQ differed across groups. In the cognitively healthy group, LDT performance and years of schooling jointly explained 66.3% of the variance in estimated IQ ($R^2 = 0.663$). | Within-population validation evidence: Supports a Brazilian Portuguese lexical-decision approach and relative stability across cognitive-status groups. The IQ criterion was contemporaneously estimated from WAIS-III subtests rather than independently observed premorbid IQ, and further validation in other Brazilian samples is needed. |
| de Oliveira et al. (2014) | 144 Brazilian adults aged 60 and older with heterogeneous educational attainment: 61 cognitively normal, 38 with amnestic MCI, and 45 with Alzheimer's disease. | WAIS-III Vocabulary as a crystallized/hold measure of premorbid IQ. | Vocabulary performance did not differ significantly across cognitively normal, amnestic MCI, and Alzheimer's disease groups, whereas fluid-intelligence measures declined with cognitive impairment. | Within-population validation evidence: Supports relative stability of Vocabulary as a hold measure in an older Brazilian sample with heterogeneous educational attainment. It does not establish cultural invariance or validity as a stand-alone Full Scale IQ estimator. |

**Table 2.** *Cont.*

| Study | Population/ Context | Premorbid Estimation Approach | Main Finding | Relevance/Limitation |
| --- | --- | --- | --- | --- |
| Yi et al. (2017) | Korean-speaking older adults, including cognitively normal, MCI, and Alzheimer's disease groups. | Korean Adult Reading Test (KART), developed around Korean orthographic-to-phonological characteristics. | KART predicted Full Scale IQ with $R^2 = 0.63$ in the validation sample; KART-predicted IQ did not differ significantly across diagnostic groups despite marked differences in current cognition. | Within-population validation evidence: Demonstrates that a reading-based estimator can be developed around Korean orthographic characteristics and retain relative stability across diagnostic groups. |
| Apolinario et al. (2013) | 132 Brazilian older adults with low educational attainment, including cognitively normal, MCI, and mild dementia groups. | Premorbid Cognitive Abilities Scale (PCAS), a structured informant-based measure of education, occupation, literacy, numeracy, technology use, and information-seeking abilities. | PCAS scores correlated with global cognition ($r = 0.73$) and reading ability ($r = 0.82$), showed high internal consistency ($\alpha = 0.90$) and inter-informant reliability (ICC = 0.96), and did not decline across diagnostic groups. | Within-population validation evidence: Supports a structured informant-based approach to estimating premorbid cognitive abilities in low-educated Brazilian older adults. Broader cultural and linguistic validation remains necessary, and the PCAS is not directly interchangeable with Full Scale IQ. |

Note. This table summarizes selected empirical studies discussed in the present narrative review that evaluated premorbid-IQ or premorbid-cognitive-ability estimators in culturally, linguistically, educationally, or migration-defined populations. Studies providing direct evidence on diversity-related influences are presented first, followed by studies providing within-population validation evidence. Studies categorized as direct evidence on diversity-related influences explicitly examined whether cultural, linguistic, educational, demographic, migration-related, or other relevant characteristics affected estimator performance, calibration, stability, or predictive validity, or directly compared estimator performance across relevant populations or contexts. Studies categorized as within-population validation evidence established the reliability, validity, or stability of an estimator within a particular language or population without necessarily examining whether diversity-related characteristics influenced estimation. Within-population validation should therefore not be interpreted as evidence that translation, adaptation, or administration in another language alone demonstrates an effect of cultural or linguistic diversity on premorbid-IQ estimation. Because this is a critical narrative review rather than a systematic or scoping review, the table is not intended to constitute an exhaustive inventory of all potentially relevant studies. Contextual studies that informed interpretation without directly evaluating a premorbid estimator are discussed in Section 4.2 rather than classified as estimator-validation evidence. HIV = human immunodeficiency virus; ICC = intraclass correlation coefficient; JART = Japanese Adult Reading Test; KART = Korean Adult Reading Test; LDT = Lexical Decision Test; MCI = mild cognitive impairment; NART-SWE = Swedish National Adult Reading Test; PCAS = Premorbid Cognitive Abilities Scale; PW = Pseudo-Words Reading; TeLPI = Irregular Word Reading Test; TOPF = Test of Premorbid Functioning; WAIS = Wechsler Adult Intelligence Scale; WAT = Word Accentuation Test; WAT-R = Word Accentuation Test-Revised; WRAT = Wide Range Achievement Test.

### *4.1. Direct Evidence on Diversity-Related Influences in Premorbid-IQ Estimation*

The direct empirical literature examining diversity-related influences on premorbid-IQ estimation remains limited. Among the clearest examples are studies that have examined whether educational background, national or regional context, language, test version, or other population characteristics alter the performance or predictive validity of reading-based or hold-type estimators. These studies are particularly informative because they address not simply whether a measure can be used in another language or population, but whether characteristics relevant to heterogeneous populations affect the resulting estimate.

Sierra Sanjurjo et al. (2015) directly examined the contribution of demographic and population characteristics to premorbid-IQ estimation in 240 adults from Argentina and

Colombia stratified by age and education. Participants completed the Word Accentuation Test-Revised (WAT-R), the WAIS-III, and a structured demographic and medical questionnaire. Education and place of birth contributed to IQ prediction, and the best-performing model incorporated WAT-R performance, years of education, and place of birth, together explaining 76% of the variance in WAIS-III Full Scale IQ. Importantly, predictive accuracy was substantially weaker at the extremes of the IQ distribution than in the average range. These findings support both the potential value of combining a language-specific reading measure with relevant demographic information and the need to evaluate whether the population and educational assumptions underlying an equation are appropriate for the individual being assessed.

Del Pino et al. (2018) directly evaluated demographic influences on Spanish-language premorbid measures in 700 healthy adults from Spain aged 18 to 86 years. Age and education significantly influenced performance on the WAT and Pseudo-Words Reading measures, whereas sex did not. Age and education jointly accounted for between 1.9% and 33.2% of variance across the premorbid-measure outcomes, and the Pseudo-Words Reading measure appeared more resistant to age-related decline than the WAT. The findings demonstrate that even within a single language and national context, premorbid reading scores cannot necessarily be interpreted independently of educational and age-related characteristics; they also provide an important null finding regarding sex.

Cross-language and normative context were examined particularly directly by Ontiveros and Gasquoine (2024). Sixty neurologically intact, balanced bilingual Mexican American adults completed both the U.S. WAIS-IV and WAIS-IV México in a counterbalanced repeated-measures design. Candidate hold measures showed different predictive relationships depending on which WAIS-IV version was being predicted, and mean scores differed systematically between the two WAIS versions. These findings demonstrate that administration in an examinee's language does not by itself establish psychometric interchangeability and that test version, language, educational history, and normative reference population may materially affect the interpretation of a candidate premorbid estimator.

Not all studies have found important subgroup differences. Olsen et al. (2015) examined the long-term stability of WRAT-3 Reading Recognition in 88 predominantly racial/ethnic minority adults with HIV over an average retest interval of approximately five years. Overall reliability was very high (ICC = 0.97), coefficients exceeded 0.92 across demographic, medical, and neurocognitive subgroups, and subgroup reliability coefficients did not differ significantly. Thus, within this clinical sample, reading stability did not appear to vary meaningfully across the subgroup characteristics examined. This finding is important because it demonstrates that diversity-related characteristics should not be assumed to undermine a reading-based estimator in every population; the relevant question is whether differential performance or instability is demonstrated empirically for the population and estimator under consideration. However, the sample was middle-aged rather than geriatric and was specific to adults with advanced HIV; accordingly, these findings are most appropriately interpreted as relevant evidence regarding subgroup stability rather than direct validation in older-adult populations.

Taken together, these studies suggest that education, population context, language and test version, and normative reference group can affect premorbid-IQ estimation in some settings, whereas other hypothesized subgroup differences are not consistently observed. The evidence therefore argues against both universal transportability and an assumption that every cultural or demographic difference requires a numerical correction. Instead, it supports explicit evaluation of the empirical applicability of a candidate estimator to the individual and population being assessed. Table 2 summarizes selected empirical studies relevant to premorbid-IQ estimation in culturally, linguistically, educationally, or

migration-defined populations and distinguishes studies providing direct evidence on diversity-related influences from those providing within-population validation evidence.

### 4.2. Empirical Contextual Evidence Relevant to Premorbid-IQ Interpretation

A broader literature provides empirical evidence regarding cultural, educational, linguistic, socioeconomic, and related characteristics that may affect the assumptions underlying premorbid-IQ estimation even when the study does not directly validate a premorbid estimator. This evidence is particularly relevant to the Contextualize component of the proposed framework because it helps identify circumstances in which common demographic variables or reading scores may not carry equivalent meaning across individuals or populations.

Educational attainment illustrates this problem particularly clearly. Manly et al. (2002) compared nondemented African American and non-Hispanic White older adults who were similar in years of formal education. African American participants initially obtained lower scores on multiple neuropsychological measures; however, after adjustment for WRAT-3 Reading performance as a proxy for educational quality, the overall racial-group effect was markedly reduced and most group differences became nonsignificant. These findings indicate that nominal years of education may not adequately represent the educational experiences relevant to later cognitive-test performance.

Subsequent work further separated educational quality from broader cultural experience. Among 503 nondemented African American adults aged 65 years and older, Manly et al. (2004) examined reading level, years of education, and acculturation simultaneously. Reading level was the strongest predictor of neuropsychological performance after accounting for age, sex, education, and acculturation. Although greater acculturation was associated with higher performance on multiple measures, its effect was substantially diminished after the other factors, particularly reading level, were considered. This pattern illustrates why broad ethnocultural categories should not be treated as direct proxies for the specific educational and experiential variables that may influence test performance.

Importantly, cross-group findings are not uniformly indicative of measurement differences. Cosentino et al. (2007) evaluated the relationship between years of education and reading performance in 342 English-speaking White, Black, and Hispanic older adults and Spanish-speaking Hispanic older adults. Reading scores increased with years of education in a comparable manner across ethnicity and language groups. Thus, although absolute scores and appropriate normative interpretation may differ across populations, the relationship between education and reading was sufficiently similar to support treating reading as a comparable construct for some theoretical and statistical purposes. This provides an important counterexample to any assumption that language or ethnicity necessarily changes the meaning of every reading-based measure.

Evidence from clinical and socioeconomically heterogeneous samples further demonstrates the limitations of treating years of schooling as equivalent to attained literacy. Ryan et al. (2005) examined 200 racially and ethnically diverse adults with HIV and found a pronounced discrepancy between reported education and reading level: approximately half of the cohort read at or below an eighth-grade level, whereas only a small proportion reported eight or fewer years of education. Reading–education discrepancy was associated with poorer neuropsychological performance, whereas racial/ethnic minority status itself was not. Dotson et al. (2009) similarly found that the relative contributions of literacy and years of education differed across race and socioeconomic strata. Literacy predicted nearly all cognitive outcomes among African American participants and lower-SES White participants, whereas education contributed little unique variance in those groups; among higher-SES White participants, both education and literacy predicted several cognitive

measures. These findings illustrate that the meaning and predictive value of educational attainment can depend on socioeconomic and historical context.

Educational quality may also influence the reading measures that are themselves commonly used as premorbid proxies. Sayegh et al. (2014) found that self-rated high-school educational quality significantly increased prediction of WRAT-4 Word Reading performance beyond race/ethnicity, socioeconomic status, and years of education. Educational quality accounted for a modest but significant increment in explained reading variance, whereas educational quality did not independently predict global neurocognitive performance after reading performance was considered. Thus, word reading may capture meaningful aspects of educational experience, but neither years of education nor self-rated educational quality should be converted directly into an informal numerical IQ adjustment.

Collectively, this contextual evidence demonstrates why premorbid-IQ estimation in heterogeneous populations requires more than recording years of education or selecting a measure in the examinee's preferred language. Literacy, educational quality, socioeconomic context, acculturation, and related experiences may affect the applicability and meaning of quantitative estimates, but the evidence does not establish fixed numerical effects that can be added to or subtracted from an estimated IQ. These findings therefore support systematic contextualization and evidentiary weighting rather than informal score correction.

### *4.3. Comparative Synthesis Across Approaches*

No single method currently has sufficient evidence to be considered universally preferable for estimating premorbid IQ across heterogeneous CALD older adults. The direct evidence on diversity-related influences summarized in Section 4.1 remains limited and is concentrated primarily in reading-based, hold, and combination approaches. By contrast, the broader contextual literature reviewed in Section 4.2 provides stronger evidence that literacy, educational experience, language and normative context, socioeconomic circumstances, and related characteristics can affect the interpretation and applicability of premorbid estimates. Within-population validation studies further demonstrate that useful estimators can be developed in multiple languages and populations, but validation within one context does not establish transportability to another. These complementary forms of evidence therefore inform different questions and should be distinguished rather than treated as interchangeable.

Within-population validation studies also provide important evidence regarding estimator applicability within particular populations. For example, Schrauf et al. (2006) found that the original Spanish Word Accentuation Test retained substantial predictive validity in older Spanish-dominant immigrants in the United States, despite concerns that adaptation might be required for this population. Such findings are clinically informative but are conceptually distinct from studies that directly test whether specific diversity-related characteristics alter estimator performance.

Across specific estimation approaches, word-reading measures have the largest body of validation evidence across multiple populations and languages, but their applicability depends on literacy, educational experience, language, writing system, normative context, and disease severity. Systematic review evidence also suggests that preservation of word-reading performance becomes less dependable with greater dementia severity (Overman et al., 2021). A measure validated in one language or population should therefore not be assumed to generalize to another simply because the underlying task is conceptually similar.

Demographic regression equations offer a different advantage because their predictors are historical and are not directly altered by current neurological impairment. However, variables such as years of education, occupation, socioeconomic status, and geographic region may have different meanings across countries, historical periods, migration contexts,

and educational systems. Combination methods may improve prediction in some settings, particularly when both demographic and performance-based components are locally validated, as demonstrated in the Latin American WAT literature (Sierra Sanjurjo et al., 2015). However, combining multiple predictors does not by itself ensure cultural validity.

Informant-based and culturally grounded approaches address limitations not fully captured by conventional psychometric methods. Structured informant measures may be particularly useful when literacy, language mismatch, or current cognitive impairment limits performance-based estimation (Apolinario et al., 2013). Culturally grounded assessment models, although not themselves premorbid-IQ estimators, illustrate how test content and task demands can be developed around culturally meaningful knowledge and everyday functioning. Machine-learning, normative-modeling, and longitudinal approaches remain promising research directions but have not yet been validated as clinical premorbid-IQ estimators in CALD older adults. Harmonization may support their future development by improving comparability across cohorts, languages, and measures, but it is not itself an estimation method.

Taken together, the available evidence supports a structured multimethod approach rather than reliance on a universally preferred estimator. The central principle is to anchor premorbid IQ estimation in the quantitative method or methods with the strongest available linguistic, educational, cultural, and psychometric applicability to the individual being assessed, while using relevant contextual information to evaluate the plausibility and relative weight of those estimates. Developmental, educational, linguistic, occupational, migration, and informant information should inform interpretation rather than generate unsupported numerical corrections. Convergence across reasonably applicable methods may strengthen confidence in an estimated premorbid range, whereas meaningful divergence should be examined and reported explicitly rather than resolved through false precision. The following section translates these principles into a structured clinical framework.

## 5. Structured Multimethod Framework for Premorbid IQ Estimation in CALD Older Adults

As summarized in the preceding sections, no currently available method consistently demonstrates strong psychometric performance while simultaneously accounting for the cultural, linguistic, educational, developmental, and related contextual heterogeneity encountered among CALD older adults. We therefore propose a structured multimethod framework organized around four clinical tasks: (1) Select the quantitative estimator or estimators with the strongest available individual applicability; (2) Contextualize each estimate using relevant developmental, educational, linguistic, occupational, migration, socioeconomic, assessment-context, and collateral information; (3) Compare reasonably applicable estimates for convergence, divergence, and identifiable sources of discrepancy; and (4) Report the most defensible premorbid range together with the relative weight assigned to each source of evidence and the degree of remaining uncertainty. The framework is intended to structure clinical reasoning rather than function as a validated scoring algorithm or generate numerical adjustments unsupported by empirical evidence. Table 3 operationalizes these four tasks.

The four tasks are directly linked to the evidence synthesized in Section 4. Select is informed primarily by evidence regarding whether a candidate estimator has adequate linguistic, educational, population, neurological, and psychometric applicability. Contextualize incorporates empirical evidence showing that years of education, literacy, educational quality, language history, socioeconomic context, migration and acculturation, and other life experiences may affect the meaning or applicability of quantitative estimates without corresponding to fixed numerical IQ adjustments. Compare requires clinicians to examine

whether multiple reasonably applicable estimates converge and, when they do not, to determine whether identifiable differences in language, literacy, normative reference group, neurological vulnerability, or other assumptions plausibly account for the discrepancy. Report makes that reasoning explicit by identifying the evidence given greatest weight, the estimated premorbid range, and remaining uncertainty. Although presented sequentially for clarity, these tasks may be iterative in practice as information obtained during contextualization or comparison leads the clinician to reconsider the applicability of an initially selected estimator.

**Table 3.** Structured Multimethod Framework and Evidence-Informed Practice Considerations for Premorbid IQ Estimation in Culturally and Linguistically Diverse Older Adults.

| Practice Consideration | Rationale | Application | Representative Evidence |
| --- | --- | --- | --- |
| SELECT: Anchor the estimate in the quantitative method or methods with the strongest individual applicability. | Demographic, hold, word-reading, and combination approaches rely on different assumptions and can fail for different reasons. | Select one or more quantitative estimators with the best available linguistic, educational, cultural, neurological, and psychometric applicability. Language-specific reading measures require validation in the target population; demographic equations require attention to development-sample and coding fit. | (Alves et al., 2012; Bright & van der Linde, 2020; Del Pino et al., 2018; Sierra Sanjurjo et al., 2015; Yi et al., 2017) |
| SELECT/CONTEXTUALIZE: Characterize language history and choose the testing language deliberately. | Language proficiency, dominance, age of acquisition, literacy, and country/language of education can alter performance on language-based hold and reading measures. | Document language acquisition and use, literacy in each language, and language and country of education. Prefer a linguistically appropriate, population-validated estimator when available; interpret cross-language discrepancies in light of test version, educational history, and normative population. | (Cosentino et al., 2007; Nielsen, 2022; Ontiveros & Gasquoine, 2024; Strutt et al., 2023) |
| SELECT/CONTEXTUALIZE: Screen for and accommodate sensory, speech, or motor barriers before interpreting performance-based estimates. | Vision, hearing, speech, and motor limitations can depress performance independently of premorbid IQ, particularly on visually presented or orally produced tasks. | Address modifiable barriers when feasible (e.g., usual corrective devices, adequate lighting, accessible presentation) and document any departure from standardized administration and its likely effect on interpretation. | (Dumassais et al., 2024; Gosselt et al., 2020; Hill-Briggs et al., 2007; Magasi et al., 2017) |
| SELECT: Use culturally grounded test-development principles without treating culturally adapted cognitive measures as premorbid-IQ estimators. | Culturally developed measures can improve ecological and cultural relevance, but a culturally appropriate cognitive screen or task does not automatically provide a validated estimate of premorbid IQ. | Use culturally grounded measures only for their validated purpose. For future premorbid measures, community-engaged development and culturally meaningful stimuli, task demands, and functional content may provide useful methodological models. | (Jacklin et al., 2020; LoGiudice et al., 2006; Sternberg et al., 2001) |
| SELECT: Use model-based approaches according to their validation status and population fit. | Regression equations depend on development-sample and coding assumptions; machine-learning and longitudinal models may capture nonlinear relationships and person-specific trajectories but depend on training data, repeated-measure properties, and external validation. | Use regression equations only when their assumptions and development sample are reasonably applicable. When longitudinal data are available, consider trajectory-based or model-based approaches as complementary research methods; do not substitute an unvalidated predictive model for an individually validated premorbid-IQ estimator. | (Barona et al., 1984; Chen et al., 2026; Marquand et al., 2016; Sierra Sanjurjo et al., 2015; Wu et al., 2022; Yuan et al., 2023) |

**Table 3.** *Cont.*

| Practice Consideration | Rationale | Application | Representative Evidence |
| --- | --- | --- | --- |
| SELECT/COMPARE: Use harmonization as research infrastructure when pooling data across tests, languages, or cohorts. | Scores from different instruments or languages are not automatically comparable, even when intended to measure the same construct. | Use appropriate linking, co-calibration, or latent-variable methods and evaluate measurement non-equivalence or differential item functioning. Harmonization can support model development and cross-study analyses but is not itself a premorbid-IQ estimator. | (Briceño et al., 2023; Merkley et al., 2023; Mukherjee et al., 2023; Vonk et al., 2022) |
| CONTEXTUALIZE: Use systematically collected contextual information to evaluate the applicability and relative weight of quantitative estimates. | Quantitative estimators may be affected by developmental, educational, linguistic, occupational, migration, cultural, and clinical factors that are not fully represented in their assumptions or normative foundations. | Use developmental, educational, linguistic, occupational, migration, and collateral information to evaluate the plausibility and relative weight of quantitative estimates. Use these data to inform interpretation rather than to generate unsupported numerical corrections. | (Bright & van der Linde, 2020; Crawford et al., 2001; Fujii, 2018; Sierra Sanjurjo et al., 2015) |
| CONTEXTUALIZE: Collect educational history beyond years completed. | Years of schooling do not capture educational quality, literacy opportunity, language of instruction, interrupted schooling, or historical inequities that may affect premorbid estimates. | Document quality and continuity of schooling, language(s) of instruction, literacy development, school interruptions, and relevant migration or sociopolitical context. Use these data to judge the applicability of demographic and reading-based estimates rather than to generate an informal IQ correction. | (Ailshire et al., 2025; Manly et al., 2002; Metcalfe et al., 2026) |
| CONTEXTUALIZE: Use structured informant information when performance-based estimation is constrained. | Low literacy, language mismatch, or current neurological impairment may reduce the validity of performance-based premorbid estimates. | When a knowledgeable informant is available, systematically document prior educational, occupational, literacy, numeracy, technological, and everyday cognitive abilities. Treat informant data according to the validated purpose and population of the specific measure rather than as a direct substitute for premorbid IQ. | (Apolinario et al., 2013) |
| CONTEXTUALIZE/REPORT: Address and document assessment-context factors and validity qualifications. | Testing language, interpreter involvement, accommodations, nonstandard procedures, evaluative anxiety, and other contextual factors can affect the validity and later interpretation of premorbid estimates. | Reduce modifiable validity threats when feasible. Report the language(s) of testing, interpreter involvement when applicable, relevant accommodations, important departures from standardized administration, and contextual factors that materially affect confidence in the estimate. | (Fujii, 2018; Nguyen et al., 2024; Régner & Huguet, 2025; Sayegh et al., 2018; Strutt et al., 2023) |
| COMPARE: Evaluate convergence, divergence, and identifiable sources of discrepancy across reasonably applicable estimates. | Different methods rely on different assumptions and may yield discrepant estimates. Convergence can strengthen confidence, whereas divergence may identify limitations in language, education, norms, neurological status, or method-specific assumptions. | Compare the most applicable estimates rather than averaging them automatically. When estimates diverge, examine plausible sources of discrepancy and adjust the relative evidentiary weight accordingly. | (Bright & van der Linde, 2020; Ontiveros & Gasquoine, 2024; Sierra Sanjurjo et al., 2015) |

**Table 3.** *Cont.*

| Practice Consideration | Rationale | Application | Representative Evidence |
| --- | --- | --- | --- |
| REPORT: Report the estimated premorbid range, evidentiary basis, and degree of uncertainty explicitly. | A single numerical estimate can imply greater precision than is warranted when available methods differ in applicability or yield discrepant results. | Identify the quantitative estimate or estimates given greatest weight and briefly state why. When reasonably applicable estimates converge, report the resulting premorbid range with corresponding confidence. When they diverge, describe the discrepancy, relevant sources of uncertainty, and the rationale for weighting particular evidence more heavily rather than forcing a single precise estimate. | (Bright & van der Linde, 2020; Crawford et al., 2001; Sierra Sanjurjo et al., 2015) |

Note. Rows are organized according to the earliest task in the proposed structured multimethod framework to which each practice consideration applies: SELECT, CONTEXTUALIZE, COMPARE, and REPORT. Practice considerations that also apply to a later task are identified with both labels. The framework and practice considerations have not been validated as a standardized clinical algorithm.

The proposed framework is complementary to, but narrower in purpose than, existing culturally informed neuropsychological guidance. Fujii's (2018) ECLECTIC framework—Education and Literacy, Culture and Acculturation, Language, Economics, Communication, Testing Situation, Intelligence Conceptualization, and Context of Immigration—provides a broad structure for identifying culturally relevant contextual factors across the neuropsychological evaluation. The present framework incorporates many of those same domains but addresses the more specific decision problem of premorbid-IQ estimation. In particular, it specifies how contextual information should be used after candidate quantitative estimators are identified: to evaluate estimator applicability and evidentiary weight, examine convergence and divergence, and communicate uncertainty rather than to generate informal numerical corrections. The Select–Contextualize–Compare–Report framework should therefore be viewed as a premorbid-IQ-specific extension that can be used within broader culturally informed approaches such as ECLECTIC, rather than as a replacement for them.

The remainder of this section applies these four tasks to common sources of interpretive uncertainty in CALD older adults. Limitations of normative representation, language and literacy, migration and socioeconomic history, trauma and testing context, sensory or motor barriers, and other clinical factors are relevant insofar as they affect selection or contextualization of an estimator; discrepancies among resulting estimates are addressed through comparison; and material limitations, accommodations, and uncertainty are carried forward into reporting.

The challenges of premorbid estimation in CALD populations extend beyond limitations of individual tests. Historically, neuropsychological research has underrepresented many ethnocultural groups, and existing normative samples often do not adequately reflect the demographic, educational, linguistic, and cultural diversity encountered in clinical practice (Arce Rentería et al., 2023; Medina et al., 2021). These limitations are further compounded by substantial within-group heterogeneity arising from differences in educational opportunities, migration experiences, socioeconomic conditions, literacy, multilingualism, and acculturation. As discussed in the preceding sections, these contextual variables may influence the validity of demographic equations, word-reading measures, and other commonly used approaches to premorbid estimation.

Trauma-informed assessment may also be relevant when estimating premorbid IQ among some CALD older adults, although trauma exposure should not be presumed on the basis of cultural, linguistic, racial, ethnic, or migration background. Trauma-informed

frameworks are well established in broader health and behavioral health settings (Substance Abuse and Mental Health Services Administration [SAMHSA], 2014), and trauma-informed approaches have also been discussed in geriatric and long-term-care contexts (O'Malley et al., 2023). Their specific application to geriatric neuropsychological assessment remains less developed. When posttraumatic stress disorder (PTSD) or clinically significant trauma-related symptoms are present, these factors may complicate interpretation of cognitive performance. Meta-analytic evidence indicates that older adults with PTSD may demonstrate lower performance across several cognitive domains, particularly learning and memory, processing speed, and executive functioning (Schuitevoerder et al., 2013). More broadly, PTSD has been associated with weaknesses in verbal learning, processing speed, attention/working memory, and verbal memory (Scott et al., 2015). However, this literature does not establish that trauma exposure, PTSD, or trauma-related hyperarousal systematically suppresses performance on single-word reading measures used to estimate premorbid IQ.

Accordingly, trauma history and current psychological symptoms are best considered as part of the broader clinical context when evaluating the validity and interpretability of premorbid estimates. Clinicians may consider whether aspects of the assessment context contribute to distress, disengagement, or reduced comfort with testing and should interpret such observations alongside the individual's broader cultural, educational, linguistic, and clinical history (Gale et al., 2023; Rivera Mindt et al., 2010). Clear communication regarding the purpose and use of test data, together with attention to culturally relevant contextual factors that may affect the assessment process, is also important (American Psychological Association [APA], 2017; Gale et al., 2023). When standardized word-reading measures are poorly matched to an individual's linguistic, educational, cultural, or clinical background, clinicians should acknowledge those limitations explicitly and integrate converging sources of information rather than treating the resulting score as an invariant estimate of premorbid ability.

These considerations also have ethical implications. The APA Ethics Code emphasizes Beneficence and Nonmaleficence, Justice, and Respect for People's Rights and Dignity, including attention to potential harm, fairness in psychological services, privacy, confidentiality, and self-determination (American Psychological Association [APA], 2017). In practice, clinicians should explain the nature and purpose of the assessment and relevant limits of confidentiality using language that is reasonably understandable to the examinee and should select and interpret assessment methods in light of evidence regarding their appropriateness for the population being assessed (American Psychological Association [APA], 2017). When the validity or reliability of a measure has not been established for the population tested, its interpretive limitations should be explicitly acknowledged. Thus, when a word-reading measure is poorly matched to an individual's language, educational history, cultural background, or clinical context, clinicians should avoid treating the resulting score as a definitive estimate of premorbid ability and instead integrate it with converging information from other relevant sources.

Age-related sensory impairments, including hearing and vision loss, as well as motor or speech limitations, may affect performance when neuropsychological tasks depend on adequate sensory input or require timed manual, written, or verbal responses. Such limitations should be considered when selecting measures and interpreting scores, because performance may partly reflect accessibility or response-output demands rather than the cognitive construct of interest (Dumassais et al., 2024; Gosselt et al., 2020; Hill-Briggs et al., 2007; Magasi et al., 2017). Failure to account for these factors may lead to underestimation of premorbid IQ, particularly when using performance-based measures such as oral word-reading tasks.

### 5.1. Example: Spanish-Language Speakers

Spanish-speaking older adults provide a useful illustration of the challenges involved in applying conventional premorbid-IQ estimation methods across CALD populations. Importantly, Spanish-speaking populations are not culturally or educationally homogeneous. Individuals may differ substantially in national origin, migration history, dialect, bilingualism, language dominance, language of education, literacy, educational quality, socioeconomic circumstances, and familiarity with standardized testing. Hispanic/Latino ethnicity should therefore not be treated as synonymous with Spanish language, nor should Spanish-speaking individuals from different countries or regions be assumed to have equivalent educational or cultural experiences (Carvajal et al., 2024; Nava et al., 2022).

These distinctions are particularly relevant when selecting intelligence measures and normative reference groups. Spanish-language versions of the Wechsler Adult Intelligence Scale (WAIS) have been developed or adapted for different Spanish-speaking populations, but use of a Spanish-language test does not by itself establish that its norms are appropriate for a particular examinee. Funes et al. (2016), for example, compared normative data from Spanish-language WAIS-III versions developed for Mexico, Spain, and Puerto Rico with the U.S. English-language WAIS-III and found meaningful differences across normative samples on identical or nearly identical subtests. The authors noted that application of Spanish-language norms could underestimate impairment in some U.S. Spanish-speaking adults, whereas application of U.S. English-language norms could overestimate impairment. Such findings illustrate that test language, normative reference group, educational background, and national or regional context must be considered jointly rather than assuming that any Spanish-language version is interchangeable with another (Funes et al., 2016). More broadly, country-specific normative studies have demonstrated meaningful effects of age, education, and other demographic characteristics on neuropsychological performance across Latin American populations (Guàrdia-Olmos et al., 2015).

Similar considerations apply to oral word-reading approaches to premorbid IQ estimation. English-language instruments such as the National Adult Reading Test and Test of Premorbid Functioning derive much of their predictive value from language-specific relationships among orthography, pronunciation, lexical knowledge, and prior educational exposure. Their underlying paradigm therefore cannot simply be translated into another language while assuming equivalent psychometric properties. Spanish-language measures such as the Word Accentuation Test (WAT) illustrate the alternative approach of developing and validating reading-based estimators around the linguistic properties of the target language. Studies conducted in Spain, Latin America, and among Spanish-speaking immigrants in the United States have demonstrated that appropriately developed Spanish-language word-reading measures can provide useful information about premorbid IQ, while also showing that their performance and predictive relationships vary with education, demographic characteristics, and population context (Del Pino et al., 2018; Del Ser et al., 1997; Schrauf et al., 2006; Sierra Sanjurjo et al., 2015). Thus, validation in one Spanish-speaking population should not automatically be generalized to another.

Reading performance itself also reflects educational and sociolinguistic experience. Cosentino et al. (2007) examined reading scores in English-speaking White, Black, and Hispanic older adults and Spanish-speaking Hispanic older adults. Although the relationship between years of education and reading performance was comparable across groups, absolute reading performance and its normative interpretation varied across ethnic and language groups. Together with evidence linking reading performance to educational quality (Sayegh et al., 2014), these findings suggest that word reading can provide meaningful information about educational experience without functioning as a culturally or linguistically context-free indicator of premorbid IQ. Accordingly, reading-based estimates

should be interpreted in relation to literacy, educational quality and continuity, language of instruction, language proficiency, and the normative population for which the measure was developed.

Substantial heterogeneity is also present within Spanish-speaking populations themselves. Buré-Reyes et al. (2013) compared adults from Chile, the Dominican Republic, Puerto Rico, and Spain and found significant group differences in delayed recall and verbal fluency despite attempts to account for demographic characteristics. Such findings caution against treating Spanish language as a sufficient proxy for cultural equivalence. Country of origin, regional language use, migration history, and exposure to Indigenous or other languages may all contribute to an individual's linguistic and educational history. Some individuals categorized as Spanish-speaking may also have acquired Spanish after another first language, while bilingual individuals may differ considerably in proficiency, dominance, age of acquisition, and patterns of language use. These characteristics may be especially important for language-mediated neuropsychological measures (Medina et al., 2023; Ontiveros & Gasquoine, 2025).

Educational context adds another layer of complexity. A recent systematic review found that measures of educational quality, most commonly literacy- or numeracy-based proxies, were generally stronger predictors of later-life neuropsychological performance than years of schooling alone, with particular relevance for populations whose educational opportunities may not be well represented by years of formal education (Metcalfe et al., 2026). Among Spanish-speaking adults living near the U.S.-Mexico border, Kamalyan et al. (2022) found that several educational, socioeconomic, and language-background characteristics were associated with cognitive performance; however, many of these associations were attenuated after applying region-specific demographic corrections, whereas bilingualism and working as a child for one's own benefit remained associated with demographically adjusted global cognition. These findings should not be interpreted as evidence that bilingualism, childhood work, migration, or other life experiences correspond to a particular premorbid-IQ level. Rather, they illustrate that contextual characteristics may affect cognitive test performance and the applicability of normative assumptions even after conventional demographic variables are considered.

The meaning of educational attainment may also differ across Hispanic/Latino heritage and migration groups. In the Study of Latinos–Investigation of Neurocognitive Aging, Meza et al. (2025) found that first-generation high school graduates of Cuban, Mexican, and Puerto Rican heritage had lower baseline cognitive scores than multigenerational high school graduates, whereas corresponding differences were not observed consistently among participants of Dominican, Central American, or South American heritage. Rates of longitudinal cognitive change were similar across groups. These findings further demonstrate that the cognitive correlates of education cannot be assumed to operate uniformly across populations with different historical, migration, and educational contexts.

Normative selection therefore remains a central concern in assessment of Spanish-speaking adults. Reviews of available U.S. Spanish-language neuropsychological norms have identified substantial progress but continued gaps in geographic, educational, and cultural representation (Morlett Paredes et al., 2021; Perez-Lao et al., 2026). A survey of U.S. neuropsychologists found that Spanish-language norms collected in the continental United States were generally preferred to foreign norms for most cognitive domains, with Mexican norms the most frequently used foreign reference data (Gasquoine et al., 2021). However, neither U.S.-based nor country-specific Spanish norms should be assumed to generalize automatically to every Spanish-speaking examinee. Studies of bilingual Mexican American adults further indicate that language proficiency can be associated with performance on language-formatted neuropsychological measures while showing little relationship with

more visually mediated measures, underscoring the importance of assessing language proficiency and test-language fit rather than relying only on self-identified bilingual status (Ontiveros & Gasquoine, 2025).

Taken together, Spanish-speaking older adults illustrate why premorbid-IQ estimation in CALD populations cannot be reduced to selecting a test in the examinee's preferred language. The strongest available quantitative estimator should be selected on the basis of its linguistic, educational, normative, and psychometric applicability to the individual. Language acquisition and dominance, literacy, language and country of education, educational quality and continuity, migration and acculturation history, occupational background, sensory functioning, and other relevant contextual information can then be used systematically to evaluate the applicability and relative weight of the resulting estimate. These contextual characteristics should not be translated into unsupported numerical IQ corrections. When no available estimator has adequate population or linguistic applicability, or when available estimates diverge substantially, the limitations and resulting uncertainty should be reported explicitly rather than resolved through false precision.

### *5.2. Brief Clinical Illustrations*

The following hypothetical examples illustrate how the preceding considerations may inform interpretation when commonly used premorbid-IQ estimators have questionable applicability to an individual's educational or linguistic background. They are intended to illustrate clinical reasoning rather than to provide scoring rules or validated algorithms.

Clinical illustration 1. An older immigrant with four years of interrupted formal schooling and limited literacy performs poorly on a word-reading estimator. Rather than treating the score as a direct indication of low premorbid IQ, the clinician examines the individual's literacy development, quality and continuity of schooling, language of instruction, occupational history, prior instrumental functioning, and collateral descriptions of prior cognitive functioning. These contextual data are used to evaluate the plausibility and limitations of the quantitative estimate rather than translated into an informal IQ correction (Fujii, 2018; Manly et al., 2002; Sayegh et al., 2014). If no quantitative estimator with adequate linguistic, educational, and psychometric applicability is available, premorbid ability should be characterized cautiously, with the basis and uncertainty of the estimate stated explicitly rather than expressed as a falsely precise IQ score.

Clinical illustration 2. A bilingual older adult who received most of their education outside the United States obtains discrepant estimates from an English word-reading measure and a language-congruent estimator validated in a population more similar to the individual's linguistic and educational background. Rather than averaging the estimates or assuming that the discrepancy represents true differences in underlying premorbid ability, the clinician considers whether language acquisition, literacy, educational history, and differences in normative reference populations may have affected performance on the English-language measure (Nielsen, 2022; Schrauf et al., 2006; Sierra Sanjurjo et al., 2015). Greater interpretive weight may be assigned to the estimate with stronger linguistic, educational, and psychometric applicability, while the discrepancy and the basis for that judgment are documented explicitly.

## 6. Summary and Practice Considerations

As summarized in Table 1, each currently available approach to estimating premorbid IQ offers important strengths but also important psychometric and cultural limitations. No single method has demonstrated adequate validity across the heterogeneous cultural, linguistic, educational, and clinical contexts represented among CALD older adults. Accordingly, selection and interpretation of premorbid IQ estimates should be guided by the

degree to which a method's assumptions and normative foundations correspond to the individual's developmental, educational, linguistic, and sociocultural background.

Although direct comparative research in CALD populations remains limited, findings from the broader premorbid-estimation literature support the potential value of considering multiple sources of information. Meyers et al. (2020), for example, found that combining multiple premorbid-estimation approaches improved prediction relative to individual methods in a heterogeneous adult sample that included several clinical and developmental-status groups. Because that study did not evaluate culturally or linguistically defined populations, however, it should be regarded as adjacent evidence for combination methods rather than evidence that multimethod estimation improves validity specifically in CALD populations. Whether similar gains occur in CALD older adults remains to be established empirically.

Accordingly, the available evidence supports a structured rather than purely impressionistic multimethod approach. Whenever feasible, estimation should be anchored in one or more quantitative methods with the strongest available linguistic, cultural, educational, and psychometric fit for the individual being evaluated. Developmental, educational, linguistic, occupational, and informant information should then be collected systematically and used to evaluate the applicability and relative weight of those quantitative estimates rather than converted into unsupported numerical corrections.

In practice, the framework can be implemented as a Select–Contextualize–Compare–Report sequence. Selection begins by determining whether language, literacy, sensory, motor, neurological, or other factors compromise a candidate estimator and then identifying the most defensible quantitative estimate or estimates available. Contextualization involves systematic consideration of relevant information, including the quality and continuity of education, literacy, language acquisition and dominance, occupational history, migration and acculturation, developmental and academic history, and collateral descriptions of prior functioning. Comparison involves examining reasonably applicable estimates for convergence, divergence, and plausible sources of discrepancy. Reporting should communicate the most defensible premorbid range, the evidence given greatest weight, and the degree of remaining uncertainty. If no available quantitative estimator has adequate linguistic, educational, cultural, and psychometric applicability, a numerical premorbid-IQ estimate should not be forced; instead, the clinician may characterize the likely premorbid range cautiously on the basis of the available evidence and explicitly document its limitations.

When multiple reasonably applicable estimates converge, a likely premorbid range may be reported with greater confidence. When estimates diverge, the discrepancy itself may be clinically informative and should be reported together with the assumptions or contextual factors that may account for it. In such cases, reporting a range or explicitly qualifying the degree of confidence is preferable to creating a falsely precise single IQ estimate.

More broadly, neuropsychological reports should document assessment conditions that materially affect interpretation, including the language or languages of testing, interpreter involvement, relevant accommodations, and material departures from standardized administration (Fujii, 2018; Nguyen et al., 2024; Sayegh et al., 2018; Strutt et al., 2023). Relevant linguistic, educational, cultural, sensory, or other contextual factors affecting confidence in a premorbid-IQ estimate should also be carried forward into the summary and diagnostic interpretation so that individual scores are not subsequently interpreted without the context in which they were obtained.

This framework is intended to make clinical reasoning more transparent and reproducible; it is not a validated scoring algorithm. Contextual variables should not be assigned numerical IQ adjustments in the absence of empirical support. Tables 3 and 4 summarize evidence-informed practice considerations and common interpretive pitfalls that follow from this structured approach.

**Table 4.** Common Interpretive Pitfalls in Premorbid IQ Estimation in Culturally and Linguistically Diverse Older Adults.

| Interpretive Pitfall | Why It Matters | Evidence-Informed Alternative | Representative Evidence |
| --- | --- | --- | --- |
| Rely on a single English word-reading score when English is not the most appropriate language for the examinee's reading assessment. | The score may reflect English literacy, language proficiency, educational history, or orthographic familiarity rather than premorbid IQ. | Use a language- and population-appropriate estimator when available and interpret it alongside educational, linguistic, and other independent evidence. | (Cosentino et al., 2007; Ontiveros & Gasquoine, 2024; Schrauf et al., 2006; Sierra Sanjurjo et al., 2015) |
| Treat years of education as equivalent across countries, eras, or educational systems. | The same number of years can represent markedly different instructional quality, literacy opportunity, language of instruction, and school continuity. | Document educational quality and continuity, literacy development, and language and country of education rather than relying on years alone. | (Ailshire et al., 2025; Manly et al., 2002; Metcalfe et al., 2026) |
| Assume that a translated, adapted, or different national version of a measure is psychometrically interchangeable. | Language, item characteristics, normative samples, and test-version differences can change score meaning and predictive relationships. | Use language-specific validation and population-appropriate norms where available; explicitly qualify interpretation when equivalence has not been demonstrated. | (Del Pino et al., 2018; Merkley et al., 2023; Ontiveros & Gasquoine, 2024) |
| Treat the single highest observed score as a definitive premorbid baseline. | Normal within-person score scatter can make the highest score an overestimate, and the selected score may itself be affected by language, education, culture, or disease. | Use best-performance information only as one source within a broader evidence-informed estimate and consider whether the supposedly preserved score is valid for the individual. | (Bright & van der Linde, 2020; Franzen et al., 1997) |
| Ignore sensory, speech, or motor limitations when interpreting performance-based estimates. | Visual, auditory, speech, or motor barriers can lower scores independently of premorbid IQ. | Address modifiable barriers when feasible, document accommodations or nonstandard administration, and qualify residual uncertainty. | (Dumassais et al., 2024; Gosselt et al., 2020; Hill-Briggs et al., 2007; Magasi et al., 2017) |
| Resolve discrepant estimates by simple averaging or unsupported numerical correction. | Disagreement among methods may identify a failed assumption (e.g., language, literacy, education, disease severity, or normative mismatch) rather than random error. | Investigate the source of discrepancy, weight estimates according to their applicability, and report a range or reduced confidence when uncertainty remains. | (Bright & van der Linde, 2020; Fujii, 2018; Sierra Sanjurjo et al., 2015) |
| Treat culturally grounded cognitive measures or structured informant measures as equivalent to a standardized premorbid-IQ score. | Culturally grounded measures may assess different constructs, and structured informant measures such as the PCAS estimate premorbid cognitive abilities rather than IQ itself. | Use each measure for its validated purpose and preserve the distinction between culturally grounded assessment, structured contextual evidence, premorbid cognitive abilities, and a validated quantitative premorbid-IQ estimate. | (Apolinario et al., 2013; Jacklin et al., 2020; LoGiudice et al., 2006) |

**Table 4.** *Cont.*

| Interpretive Pitfall | Why It Matters | Evidence-Informed Alternative | Representative Evidence |
| --- | --- | --- | --- |
| Interpret hesitation, anxiety, unfamiliarity with testing, or discomfort with oral reading as cognitive weakness without considering context. | Evaluation context, stereotype threat, cultural expectations, and prior experiences can influence observed performance independently of underlying ability. | Interpret behavior in context, reduce unnecessary evaluative pressure without altering standardized task demands, and use alternative indicators when task validity is materially compromised. | (Ardila, 2005; Fujii, 2018; Régner & Huguet, 2025) |
| Treat machine-learning, longitudinal modeling, or harmonization as established clinical substitutes for validated premorbid-IQ estimation. | Predictive accuracy can conceal subgroup error; longitudinal data can be distorted by practice, attrition, cohort, and measurement effects; and harmonization is methodological infrastructure rather than an individual estimator. | Use emerging models as research complements until they demonstrate appropriate external validation, subgroup calibration, measurement equivalence, and clinical interpretability for the target population. Use harmonization to support comparability and model development, not as an individual premorbid-IQ estimate. | (Briceño et al., 2023; Chen et al., 2026; Mukherjee et al., 2023; Vonk et al., 2022; Wu et al., 2022; Yuan et al., 2023) |

Note. These interpretive pitfalls and evidence-informed alternatives synthesize the empirical and conceptual literature reviewed in this article but have not been validated as a standardized clinical algorithm. PCAS = Premorbid Cognitive Abilities Scale.

## 7. Limitations

This review has several limitations. First, it is a critical narrative review rather than a systematic or scoping review. The literature searches were designed to support broad conceptual and clinical synthesis and therefore cannot guarantee identification of every relevant publication. Relatedly, we did not conduct a formal risk-of-bias assessment, systematic evidence-grading procedure, or quantitative meta-analysis. Accordingly, the evidence summaries and practice considerations presented in this review should be interpreted as a narrative synthesis rather than as formal comparative evidence grades or a ranked hierarchy of premorbid-IQ estimation methods.

Second, direct empirical evidence on diversity-related influences in premorbid-IQ estimation among CALD older adults remains limited. The review therefore also draws on within-population validation studies and contextual evidence from broader adult samples and from research on literacy, educational quality, cross-cultural neuropsychology, measurement equivalence, and related methodological issues. Within-population validation evidence can establish the performance of an estimator within a particular population but does not by itself demonstrate the effects of diversity-related characteristics or establish transportability to another context. Contextual evidence can inform interpretation and identify important sources of potential bias but does not constitute direct validation of a premorbid-IQ estimator.

Third, the available literature is highly heterogeneous with respect to populations, languages, educational backgrounds, instruments, clinical conditions, and validity criteria. This heterogeneity, together with the narrative nature of the review, limits direct comparison across studies and does not support formal quantitative ranking of estimation approaches.

Finally, the structured clinical framework proposed here represents an evidence-informed synthesis but has not itself been empirically validated as a standardized clinical algorithm. The framework and practice considerations proposed in this review should

therefore be regarded as provisional clinical considerations and hypotheses for further study rather than as an established scoring system or universally validated assessment protocol.

## 8. Future Research

As populations age and become increasingly CALD, stronger empirical evidence is needed to guide premorbid IQ estimation in CALD older adults. Key unanswered questions include which existing premorbid-IQ estimators perform most accurately across specific CALD populations; whether combining methods improves accuracy beyond the best-fitting individual estimator; how educational quality, literacy, language history, migration, and other contextual variables should be incorporated without introducing unsupported adjustments; and whether newer longitudinal, computational, or culturally grounded approaches provide meaningful incremental validity over established methods.

Future studies should directly compare demographic, reading-based, informant-based, and combination approaches within the same well-characterized CALD samples rather than inferring their relative performance across unrelated studies. Prospective studies should also compare estimated premorbid IQ with actual prior intellectual, educational, or longitudinal cognitive data whenever such information is available. Particular attention is needed to older adults with low literacy, interrupted or limited formal schooling, multilingual backgrounds, and education obtained outside dominant Western educational systems.

Additional research is needed to determine how contextual information should be collected and incorporated into clinical interpretation. Standardized procedures for documenting educational quality, literacy, language history, migration and acculturation, occupational complexity, and collateral reports of prior functioning could improve reproducibility and permit examination of interrater agreement. Culturally and linguistically grounded measures should also undergo explicit evaluation of measurement equivalence and external validation across regions, dialects, migration groups, and educational strata. Emerging computational and longitudinal approaches should similarly be evaluated for prediction error, calibration, transportability, and subgroup performance rather than predictive accuracy alone.

Community-based participatory approaches may further strengthen this work by involving CALD older adults, care partners, clinicians, community stakeholders, and professional interpreters in identifying culturally meaningful constructs, assessment procedures, and implementation barriers. While the empirical evidence base develops, structured expert-consensus approaches may also help identify areas of provisional agreement for clinical practice. One such approach is the Delphi method, a systematic consensus-building procedure in which a panel of experts completes successive rounds of ratings or questionnaires, receives controlled feedback regarding the group's responses, and has opportunities to reconsider its ratings across rounds (Dalkey & Helmer, 1963; Jorm, 2015). The method is particularly useful when empirical evidence is incomplete and the goal is to identify areas of expert agreement while limiting some of the interpersonal pressures inherent in face-to-face consensus processes (Jorm, 2015). A Delphi study involving neuropsychologists with expertise in cultural neuropsychology, aging, psychometrics, and multilingual assessment could therefore provide provisional guidance regarding assessment and reporting practices while highlighting priorities requiring direct empirical validation.

**Author Contributions:** Conceptualization, P.S. (Philip Sayegh), L.D.M., C.Y.M.P., P.S. (Paola Suarez) and C.M.N.; methodology, P.S. (Philip Sayegh), L.D.M., C.Y.M.P., P.S. (Paola Suarez) and C.M.N.; investigation, P.S. (Philip Sayegh), L.D.M., C.Y.M.P., P.S. (Paola Suarez) and C.M.N.; writing—original draft preparation, P.S. (Philip Sayegh), L.D.M., C.Y.M.P., P.S. (Paola Suarez) and C.M.N.; writing—review and editing, P.S. (Philip Sayegh), L.D.M., C.Y.M.P., P.S. (Paola Suarez) and C.M.N.; project administration, P.S. (Philip Sayegh). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

Abdelsalam, S. M., Matter, A. H. M., El-Hariri, H. M., Assaf, A. H., Khater, M. S., & Tawfik, H. M. (2025). Adaptation and validation of the modified Egyptian Arabic version of Addenbrooke's cognitive examination III (VI-ACE-III) for assessing cognitive impairment in visually impaired elderly. *BMC Geriatrics*, *25*(1), 145. [CrossRef] [PubMed]

Abdullah, L. B., Khandakar, I., Douglas, A., Nance, R., Zhou, Z., Hall, J., O'Bryant, S., & HABS-HD Study Team. (2026). Predicting low premorbid cognitive ability with social determinants: A machine learning approach. *The Journal of Aging Research & Lifestyle*, *15*, 100062. [CrossRef] [PubMed]

Administration for Community Living (ACL). (2024). *2023 profile of older Americans*. U.S. Department of Health and Human Services. Available online: https://acl.gov/sites/default/files/Profile%20of%20OA/ACL_ProfileOlderAmericans2023_508.pdf (accessed on 6 November 2025).

Ailshire, J. A., Farina, M. P., Jackson, H., & Walsemann, K. M. (2025). The role of educational attainment and quality in U.S. regional variation in prevalence of dementia and CIND. *PLoS ONE*, *20*(9), e0332410. [CrossRef] [PubMed]

Alves, L., Simões, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: The irregular word reading test (TeLPI). *Archives of Clinical Neuropsychology*, *27*(1), 58–68. [CrossRef] [PubMed]

Alves, L., Simões, M. R., Martins, C., Freitas, S., & Santana, I. (2013). TeLPI performance in subjects with mild cognitive impairment and Alzheimer disease: A validation study. *Alzheimer Disease & Associated Disorders*, *27*(4), 324–329. [CrossRef] [PubMed]

American Psychological Association (APA). (2017). *Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017)*. Available online: https://www.apa.org/ethics/code (accessed on 14 September 2026).

Apolinario, D., Brucki, S. M. D., Ferretti, R. E. L., Farfel, J. M., Magaldi, R. M., Busse, A. L., & Jacob-Filho, W. (2013). Estimating premorbid cognitive abilities in low-educated populations. *PLoS ONE*, *8*(3), e60084. [CrossRef] [PubMed]

Arce Rentería, M., Mobley, T. M., Evangelista, N. D., Medina, L. D., Deters, K. D., Fox-Fuller, J. T., Minto, L. R., Avila-Rieger, J., & Bettcher, B. M. (2023). Representativeness of samples enrolled in Alzheimer's disease research centers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *15*(2), e12450. [CrossRef] [PubMed]

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review*, *15*(4), 185–195. [CrossRef] [PubMed]

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS–R. *Journal of Consulting and Clinical Psychology*, *52*(5), 885–887. [CrossRef]

Bertola, L., Benseñor, I. M., Gross, A. L., Caramelli, P., Barreto, S. M., Moreno, A. B., Griep, R. H., Viana, M. C., Lotufo, P. A., & Suemoto, C. K. (2021). Longitudinal measurement invariance of neuropsychological tests in a diverse sample from the ELSA-Brasil study. *Brazilian Journal of Psychiatry*, *43*(3), 254–261. [CrossRef] [PubMed]

Briceño, E. M., Arce Rentería, M., Gross, A. L., Jones, R. N., Gonzalez, C., Wong, R., Weir, D. R., Langa, K. M., & Manly, J. J. (2023). A cultural neuropsychological approach to harmonization of cognitive data across culturally and linguistically diverse older adult populations. *Neuropsychology*, *37*(3), 247–257. [CrossRef] [PubMed]

Brickman, A. M., Cabo, R., & Manly, J. J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology*, *13*(2), 91–100. [CrossRef] [PubMed]

Bright, P., & van der Linde, I. (2020). Comparison of methods for estimating premorbid intelligence. *Neuropsychological Rehabilitation*, *30*(1), 1–14. [CrossRef] [PubMed]

Buré-Reyes, A., Hidalgo-Ruzzante, N., Vilar-López, R., Gontier, J., Sánchez, L., Pérez-García, M., & Puente, A. E. (2013). Neuropsychological test performance of Spanish speakers: Is performance different across different Spanish-speaking subgroups? *Journal of Clinical and Experimental Neuropsychology*, *35*(4), 404–412. [CrossRef] [PubMed]

Burin, D. I., Jorge, R. E., Arizaga, R. A., & Paulsen, J. S. (2000). Estimation of premorbid intelligence: The word accentuation test–Buenos Aires version. *Journal of Clinical and Experimental Neuropsychology*, *22*(5), 677–685. [CrossRef] [PubMed]

Byrd, D. A., & Rivera Mindt, M. G. (2022). Neuropsychology's race problem does not begin or end with demographically adjusted norms. *Nature Reviews Neurology*, *18*(3), 125–126. [CrossRef] [PubMed]

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. [CrossRef] [PubMed]

Carvajal, D. N., Anaya, Y. B., McLean, I., Aragón, M., Figueroa, E., Plasencia, G., Martinez-Bianchi, V., & Rodríguez, J. E. (2024). We are not all the same: Implications of heterogeneity among Latiné/e/x/o/a, Hispanic, and Spanish origin people. *Annals of Family Medicine*, *22*(3), 254–258. [CrossRef] [PubMed]

Chen, Y., Kong, S., Wang, T., Chen, J., Ma, K., Zhang, J., Zhang, Y., & Wang, M. (2026). From symptom tracking to prevention—A transformer-based dynamic model for predicting mild cognitive impairment risk in older adults with depression: A longitudinal study based on CHARLS and CLHLS. *Medicine*, *105*(19), e48653. [CrossRef] [PubMed]

Cosentino, S., Manly, J., & Mungas, D. (2007). Do reading tests measure the same construct in multiethnic and multilingual older persons? *Journal of the International Neuropsychological Society*, *13*(2), 228–236. [CrossRef] [PubMed]

Crawford, J. R., Millar, J., & Milne, A. B. (2001). Estimating premorbid IQ from demographic variables: A comparison of a regression equation vs. clinical judgement. *British Journal of Clinical Psychology*, *40*(1), 97–105. [CrossRef] [PubMed]

Crawford, J. R., Stewart, L. E., Cochrane, R. H. B., Foulds, J. A., Besson, J. A. O., & Parker, D. M. (1989). Estimating premorbid IQ from demographic variables: Regression equations derived from a UK sample. *British Journal of Clinical Psychology*, *28*(3), 275–278. [CrossRef] [PubMed]

Crawford, J. R., Venneri, A., & O'Carroll, R. E. (1998). Neuropsychological assessment of the elderly. In B. A. Edelstein (Ed.), *Comprehensive clinical psychology: Vol. 7. Clinical geropsychology* (pp. 133–169). Pergamon. [CrossRef]

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458–467. [CrossRef]

delCacho-Tena, A., Christ, B. R., Arango-Lasprilla, J. C., Perrin, P. B., Rivera, D., & Olabarrieta-Landa, L. (2024). Normative data estimation in neuropsychological tests: A systematic review. *Archives of Clinical Neuropsychology*, *39*(3), 383–398. [CrossRef] [PubMed]

Del Pino, R., Peña, J., Ibarretxe-Bilbao, N., Schretlen, D. J., & Ojeda, N. (2018). Demographically calibrated norms for two premorbid intelligence measures: The word accentuation test and pseudo-words reading subtest. *Frontiers in Psychology*, *9*, 1950. [CrossRef] [PubMed]

Del Ser, T., González-Montalvo, J. I., Martínez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the word accentuation test and its application to the diagnosis of dementia. *Brain and Cognition*, *33*(3), 343–356. [CrossRef] [PubMed]

de Oliveira, M. O., Nitrini, R., Yassuda, M. S., & Brucki, S. M. D. (2014). Vocabulary is an appropriate measure of premorbid intelligence in a sample with heterogeneous educational level in Brazil. *Behavioural Neurology*, *2014*, 875960. [CrossRef] [PubMed]

Dotson, V. M., Kitner-Triolo, M. H., Evans, M. K., & Zonderman, A. B. (2009). Effects of race and socioeconomic status on the relative influence of education and literacy on cognitive functioning. *Journal of the International Neuropsychological Society*, *15*(4), 580–589. [CrossRef] [PubMed]

Dumassais, S., Pichora-Fuller, M. K., Guthrie, D., Phillips, N. A., Savundranayagam, M., & Wittich, W. (2024). Strategies used during the cognitive evaluation of older adults with dual sensory impairment: A scoping review. *Age and Ageing*, *53*(3), afae051. [CrossRef] [PubMed]

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology*, *12*(8), 711–738. [CrossRef]

Fujii, D. E. M. (2018). Developing a cultural context for conducting a neuropsychological evaluation with a culturally diverse client: The ECLECTIC framework. *The Clinical Neuropsychologist*, *32*(8), 1356–1392. [CrossRef] [PubMed]

Funes, C. M., Hernandez Rodriguez, J., & Lopez, S. R. (2016). Norm comparisons of the Spanish-language and English-language WAIS-III: Implications for clinical assessment and test adaptation. *Psychological Assessment*, *28*(12), 1709–1715. [CrossRef] [PubMed]

Gale, R., Fellner, K., Tomlinson, G., & Danto, D. (2023). Beyond appropriate norms: Cultural safety with Indigenous people in Canadian neuropsychology. *The Journal of Concurrent Disorders*, *5*(1), 52–79. [CrossRef]

Gasquoine, P. G. (2009). Race-norming of neuropsychological tests. *Neuropsychology Review*, *19*(2), 250–262. [CrossRef] [PubMed]

Gasquoine, P. G., Weimer, A. A., Estevis, E., & Perez, E. (2021). Survey of Spanish language neuropsychological test use in the assessment of Hispanic Americans/Latino/as/x. *Archives of Clinical Neuropsychology*, *36*(7), 1350–1360. [CrossRef] [PubMed]

Gerstorf, D., Ram, N., Hoppmann, C., Willis, S. L., & Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology*, *47*(4), 1026–1041. [CrossRef] [PubMed]

Gosselt, I. K., Nijboer, T. C. W., & Van Es, M. A. (2020). An overview of screening instruments for cognition and behavior in patients with ALS: Selecting the appropriate tool for clinical practice. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *21*(5–6), 324–336. [CrossRef] [PubMed]

Guàrdia-Olmos, J., Peró-Cebollero, M., Rivera, D., & Arango-Lasprilla, J. C. (2015). Methodology for the development of normative data for ten Spanish-language neuropsychological tests in eleven Latin American countries. *NeuroRehabilitation*, *37*(4), 493–499. [CrossRef] [PubMed]

Hawkins, K. A. (1995). Limitations to the validity of the Barona regression formula and similar demographically-based methods of estimating preinjury intellectual functioning. *Behavioral Sciences & the Law*, *13*(4), 491–503. [CrossRef]

Hill-Briggs, F., Dial, J. G., Morere, D. A., & Joyce, A. (2007). Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness. *Archives of Clinical Neuropsychology*, *22*(3), 389–404. [CrossRef] [PubMed]

Jacklin, K., Pitawanakwat, K., Blind, M., O'Connell, M. E., Walker, J., Lemieux, A. M., & Warry, W. (2020). Developing the Canadian indigenous cognitive assessment for use with Indigenous older Anishinaabe adults in Ontario, Canada. *Innovation in Aging*, *4*(4), igaa038. [CrossRef] [PubMed]

Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry*, *49*(10), 887–897. [CrossRef] [PubMed]

Judd, T., Capetillo, D., Carrión-Baralt, J., Mármol, L. M., San Miguel-Montes, L., Navarrete, M. G., Puente, A. E., Rodas Romero, H., Valdés, J., & NAN Policy and Planning Committee. (2009). Professional considerations for improving the neuropsychological evaluation of Hispanics: A national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology*, *24*(2), 127–135. [CrossRef] [PubMed]

Kamalyan, L., Guareña, L. A., Díaz-Santos, M., Suarez, P., Cherner, M., García Alcorn, M. Y., Umlauf, A., Franklin, D. R., Rivera Mindt, M., Artiola i Fortuny, L., Heaton, R. K., & Marquine, M. J. (2022). Influence of educational background, childhood socioeconomic environment, and language use on cognition among Spanish-speaking Latinos living near the US–Mexico border. *Journal of the International Neuropsychological Society*, *28*(8), 876–890. [CrossRef] [PubMed]

Kareken, D. A. (1997). Judgment pitfalls in estimating premorbid intellectual function. *Archives of Clinical Neuropsychology*, *12*(8), 701–709. [CrossRef]

Kareken, D. A., & Williams, J. M. (1994). Human judgment and estimation of premorbid intellectual function. *Psychological Assessment*, *6*(2), 83–91. [CrossRef]

Kirton, J. W., Soble, J. R., Marceaux, J. C., Messerly, J., Bain, K. M., Webber, T. A., Fullen, C., Alverson, W. A., & McCoy, K. J. M. (2020). Comparison of models of premorbid IQ estimation using the TOPF, OPIE-3, and Barona equation, with corrections for the Flynn effect. *Neuropsychology*, *34*(1), 43–52. [CrossRef] [PubMed]

Kiselica, A. M., Karr, J. E., Mikula, C. M., Ranum, R. M., Benge, J. F., Medina, L. D., & Woods, S. P. (2024). Recent advances in neuropsychological test interpretation for clinical practice. *Neuropsychology Review*, *34*(2), 637–667. [CrossRef] [PubMed]

Krueger, K. R., Lam, C. S., & Wilson, R. S. (2006). The word accentuation test–Chicago. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1201–1207. [CrossRef] [PubMed]

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist*, *9*(1), 83–88. [CrossRef]

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). Oxford University Press.

LoGiudice, D., Smith, K., Thomas, J., Lautenschlager, N. T., Almeida, O. P., Atkinson, D., & Flicker, L. (2006). Kimberley indigenous cognitive assessment tool (KICA): Development of a cognitive assessment tool for older indigenous Australians. *International Psychogeriatrics*, *18*(2), 269–280. [CrossRef] [PubMed]

Magasi, S., Harniss, M., Tulsky, D. S., Cohen, M. L., Heaton, R. K., & Heinemann, A. W. (2017). Test accommodations for individuals with neurological conditions completing the NIH toolbox–cognition battery: An evaluation of frequency and appropriateness. *Rehabilitation Psychology*, *62*(4), 455–463. [CrossRef] [PubMed]

Manly, J. J. (2008). Critical issues in cultural neuropsychology: Profit from diversity. *Neuropsychology Review*, *18*(3), 179–183. [CrossRef] [PubMed]

Manly, J. J., Byrd, D. A., Touradji, P., & Stern, Y. (2004). Acculturation, reading level, and neuropsychological test performance among African American elders. *Applied Neuropsychology*, *11*(1), 37–46. [CrossRef] [PubMed]

Manly, J. J., & Echemendia, R. J. (2007). Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Archives of Clinical Neuropsychology*, *22*(3), 319–325. [CrossRef] [PubMed]

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*(3), 341–348. [CrossRef] [PubMed]

Marier, A., Dadar, M., Bouhali, F., Montembeault, M., & Alzheimer's Disease Neuroimaging Initiative. (2024). Irregular word reading as a marker of semantic decline in Alzheimer's disease: Implications for premorbid intellectual ability measurement. *Alzheimer's Research & Therapy*, *16*, 96. [CrossRef] [PubMed]

Marquand, A. F., Rezek, I., Buitelaar, J., & Beckmann, C. F. (2016). Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biological Psychiatry*, *80*(7), 552–561. [CrossRef] [PubMed]

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of national adult reading test. *Psychiatry and Clinical Neurosciences*, *60*(3), 332–339. [CrossRef] [PubMed]

Medina, L. D., Henry, S., Torres, S., MacDonald, B., & Strutt, A. M. (2023). The measurement of acculturation in neuropsychological evaluations of Hispanic/Latino individuals across the lifespan: A scoping review of the literature. *Archives of Clinical Neuropsychology*, *38*(3), 365–386. [CrossRef] [PubMed]

Medina, L. D., Torres, S., Gioia, A., Ochoa Lopez, A., Wang, J., & Cirino, P. T. (2021). Reporting of demographic variables in neuropsychological research: An update of O'Bryant et al.'s trends in the current literature. *Journal of the International Neuropsychological Society*, *27*(5), 497–507. [CrossRef] [PubMed]

Mendoza, L., Garcia, P., Duara, R., Rosselli, M., Loewenstein, D., Greig-Custo, M. T., Barker, W., Dahlin, P., & Rodriguez, M. J. (2022). The effect of acculturation on cognitive performance among older Hispanics in the United States. *Applied Neuropsychology: Adult*, *29*(2), 163–171. [CrossRef] [PubMed]

Merkley, T. L., Esopenko, C., Zizak, V. S., Bilder, R. M., Strutt, A. M., Tate, D. F., & Irimia, A. (2023). Challenges and opportunities for harmonization of cross-cultural neuropsychological data. *Neuropsychology*, *37*(3), 237–246. [CrossRef] [PubMed]

Metcalfe, T., Nielsen, T. R., O'Donald, F., Franzen, S., & Calia, C. (2026). A systematic review of the measurement and influence of quality of education on neuropsychological test performance in later life. *Neuropsychology Review*. Advance online publication. [CrossRef] [PubMed]

Meyers, J. E., Miller, R. M., Rohling, M. L., & Kalat, S. S. (2020). Premorbid estimates of neuropsychological functioning for diverse groups. *Applied Neuropsychology: Adult*, *27*(4), 364–375. [CrossRef] [PubMed]

Meza, E., Tarraf, W., Gallo, L. C., Isasi, C. R., Perreira, K. M., Lamar, M., Estrella, M. L., Daviglus, M., Allen, I. E., Glymour, M. M., Torres, J. M., & González, H. M. (2025). Intergenerational upward educational mobility and cognitive performance: Results from the study of Latinos–investigation of neurocognitive aging (SOL-INCA). *Alzheimer's & Dementia*, *21*(5), e70214. [CrossRef] [PubMed]

Morlett Paredes, A., Gooding, A., Artiola i Fortuny, L., Rivera Mindt, M., Suárez, P., Scott, T. M., Heaton, A., Heaton, R. K., Cherner, M., & Marquine, M. J. (2021). The state of neuropsychological test norms for Spanish-speaking adults in the United States. *The Clinical Neuropsychologist*, *35*(2), 236–252. [CrossRef] [PubMed]

Mortensen, E. L., Gade, A., & Reinisch, J. M. (1991). A critical note on Lezak's "best performance method" in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *13*(2), 361–371. [CrossRef] [PubMed]

Mukherjee, S., Choi, S.-E., Lee, M. L., Scollard, P., Trittschuh, E. H., Mez, J., Saykin, A. J., Gibbons, L. E., Sanders, R. E., Zaman, A. F., Teylan, M. A., Kukull, W. A., Barnes, L. L., Bennett, D. A., LaCroix, A. Z., Larson, E. B., Cuccaro, M., Mercado, S., Dumitrescu, L., ... Crane, P. K. (2023). Cognitive domain harmonization and cocalibration in studies of older adults. *Neuropsychology*, *37*(4), 409–423. [CrossRef] [PubMed]

Nava, A., Estrada, L., Gerchow, L., Scott, J., Thompson, R., & Squires, A. (2022). Grouping people by language exacerbates health inequities—The case of Latinx/Hispanic populations in the US. *Research in Nursing & Health*, *45*(2), 142–147. [CrossRef] [PubMed]

NCS Pearson. (2009). *Advanced clinical solutions for WAIS-IV and WMS-IV: Administration and scoring manual*. The Psychological Corporation.

Nelson, H. E. (1982). *National adult reading test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*. NFER-Nelson.

Nguyen, C. M., Rampa, S., Staios, M., Nielsen, T. R., Zapparoli, B., Zhou, X. E., Mbakile-Mahlanza, L., Colon, J., Hammond, A., Hendriks, M., Kgolo, T., Serrano, Y., Marquine, M. J., Dutt, A., Evans, J., & Judd, T. (2024). Neuropsychological application of the international test commission guidelines for translation and adapting of tests. *Journal of the International Neuropsychological Society*, *30*(7), 621–634. [CrossRef] [PubMed]

Nielsen, T. R. (2022). Cognitive assessment in culturally, linguistically, and educationally diverse older populations in Europe. *American Journal of Alzheimer's Disease & Other Dementias*, *37*, 15333175221117006. [CrossRef] [PubMed]

Nielsen, T. R., Franzen, S., Watermeyer, T., Jiang, J., Calia, C., Kjærgaard, D., Bothe, S., & Mukadam, N. (2024). Interpreter-mediated neuropsychological assessment: Clinical considerations and recommendations from the European Consortium on Cross-Cultural Neuropsychology (ECCroN). *The Clinical Neuropsychologist*, *38*(8), 1775–1805. [CrossRef] [PubMed]

Nogueira, J., Gerardo, B., Santana, I., Simões, M. R., & Freitas, S. (2022). The assessment of cognitive reserve: A systematic review of the most used quantitative measurement methods of cognitive reserve for aging. *Frontiers in Psychology*, *13*, 847186. [CrossRef] [PubMed]

Olsen, J. P., Fellows, R. P., Rivera-Mindt, M., Morgello, S., Byrd, D. A., & Manhattan HIV Brain Bank. (2015). Reading ability as an estimator of premorbid intelligence: Does it remain stable among ethnically diverse HIV+ adults? *The Clinical Neuropsychologist*, *29*(7), 1034–1052. [CrossRef] [PubMed]

O'Malley, K. A., Sullivan, J. L., Mills, W., Driver, J., & Moye, J. (2023). Trauma-informed care in long-term care settings: From policy to practice. *The Gerontologist*, *63*(5), 803–811. [CrossRef] [PubMed]

Ontiveros, G., & Gasquoine, P. G. (2024). Wechsler adult intelligence scale–IV México versus U.S. versions in the assessment of Mexican Americans. *Journal of the International Neuropsychological Society*, *30*(9), 819–827. [CrossRef] [PubMed]

Ontiveros, G., & Gasquoine, P. G. (2025). Clinical neuropsychology of bilingual Mexican American adults: Effect of language proficiency and dominance. *Archives of Clinical Neuropsychology*, *40*(3), 456–466. [CrossRef] [PubMed]

Overman, M. J., Leeworthy, S., & Welsh, T. J. (2021). Estimating premorbid intelligence in people living with dementia: A systematic review. *International Psychogeriatrics*, *33*(11), 1145–1159. [CrossRef] [PubMed]

Perez-Lao, A., Ying, G., Arias, F., Levy, S. A., & Smith, G. E. (2026). A systematic review of available normative data on neuropsychological tests for Spanish speakers in the U.S., Latin America and the Caribbean, and Spain. *Neuropsychology Review*, *36*(2), 113–139. [CrossRef] [PubMed]

Possin, K. L., Tsoy, E., & Windon, C. C. (2021). Perils of race-based norms in cognitive testing: The case of former NFL players. *JAMA Neurology*, *78*(4), 377–378. [CrossRef] [PubMed]

Powell, B. D., Brossart, D. F., & Reynolds, C. R. (2003). Evaluation of the accuracy of two regression-based methods for estimating premorbid IQ. *Archives of Clinical Neuropsychology*, *18*(3), 277–292. [CrossRef]

Reynolds, C. R. (1997). Postscripts on premorbid ability estimation: Conceptual addenda and a few words on alternative and conditional approaches. *Archives of Clinical Neuropsychology*, *12*(8), 769–778. [CrossRef]

Régner, I., & Huguet, P. (2025). Age-based stereotype threat effects: From the laboratory to the clinical setting. *Cortex*, *183*, 261–273. [CrossRef] [PubMed]

Rivera, D., Forte, A., Olabarrieta-Landa, L., Perrin, P. B., & Arango-Lasprilla, J. C. (2024). Methodology for the generation of normative data for the U.S. adult Spanish-speaking population: A Bayesian approach. *NeuroRehabilitation*, *55*(2), 155–167. [CrossRef] [PubMed]

Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. *The Clinical Neuropsychologist*, *24*(3), 429–453. [CrossRef] [PubMed]

Rolstad, S., Nordlund, A., Gustavsson, M. H., Eckerström, C., Klang, O., Hansen, S., & Wallin, A. (2008). The Swedish national adult reading test (NART-SWE): A test of premorbid IQ. *Scandinavian Journal of Psychology*, *49*(6), 577–582. [CrossRef] [PubMed]

Ryan, E. L., Baird, R., Rivera Mindt, M., Byrd, D., Monzones, J., Morgello, S., & the Manhattan HIV Brain Bank. (2005). Neuropsychological impairment in racial/ethnic minorities with HIV infection and low literacy levels: Effects of education and reading level in participant characterization. *Journal of the International Neuropsychological Society*, *11*(7), 889–898. [CrossRef] [PubMed]

Sayegh, P., Arentoft, A., Thaler, N. S., Dean, A. C., & Thames, A. D. (2014). Quality of education predicts performance on the wide range achievement test-4th edition word reading subtest. *Archives of Clinical Neuropsychology*, *29*(8), 731–736. [CrossRef] [PubMed]

Sayegh, P., Funes, C. M., & Suarez, P. (2018). Dementia assessment in Latino Americans. In G. Yeo, L. A. Gerdner, & D. Gallagher-Thompson (Eds.), *Ethnicity and the dementias* (3rd ed., pp. 143–162). Routledge. [CrossRef]

Schoenberg, M. R., Scott, J. G., Duff, K., & Adams, R. L. (2002). Estimation of WAIS-III intelligence from combined performance and demographic variables: Development of the OPIE-3. *The Clinical Neuropsychologist*, *16*(4), 426–437. [CrossRef] [PubMed]

Schrauf, R. W., Weintraub, S., & Navarro, E. (2006). Is adaptation of the word accentuation test of premorbid intelligence necessary for use among older, Spanish-speaking immigrants in the United States? *Journal of the International Neuropsychological Society*, *12*(3), 391–399. [CrossRef] [PubMed]

Schuitevoerder, S., Rosen, J. W., Twamley, E. W., Ayers, C. R., Sones, H., Lohr, J. B., Goetter, E. M., Fonzo, G. A., Holloway, K. J., & Thorp, S. R. (2013). A meta-analysis of cognitive functioning in older adults with PTSD. *Journal of Anxiety Disorders*, *27*(6), 550–558. [CrossRef] [PubMed]

Scott, J. C., Matt, G. E., Wrocklage, K. M., Crnich, C., Jordan, J., Southwick, S. M., Krystal, J. H., & Schweinsburg, B. C. (2015). A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychological Bulletin*, *141*(1), 105–140. [CrossRef] [PubMed]

Serrao, V. T., Brucki, S. M. D., Campanholo, K. R., Mansur, L. L., Nitrini, R., & Miotto, E. C. (2015). Performance of a sample of patients with mild cognitive impairment (MCI), Alzheimer's disease (AD) and healthy elderly on a lexical decision test (LDT) as a measure of pre-morbid intelligence. *Dementia & Neuropsychologia*, *9*(3), 265–269. [CrossRef] [PubMed]

Shura, R. D., Ord, A. S., Martindale, S. L., Miskey, H. M., & Taber, K. H. (2022). Test of premorbid functioning: You're doing it wrong, but does it matter? *Archives of Clinical Neuropsychology*, *37*(5), 1035–1040. [CrossRef] [PubMed]

Sierra Sanjurjo, N., Montañes, P., Sierra Matamoros, F. A., & Burin, D. (2015). Estimating intelligence in Spanish: Regression equations with the word accentuation test and demographic variables in Latin America. *Applied Neuropsychology: Adult*, *22*(4), 252–261. [CrossRef] [PubMed]

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). Oxford University Press.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028. [CrossRef] [PubMed]

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., Vuoksimaa, E., & The

Reserve, Resilience and Protective Factors PIA Empirical Definitions and Conceptual Frameworks Workgroup. (2020). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, *16*(9), 1305–1311. [CrossRef] [PubMed]

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*(5), 401–418. [CrossRef]

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Strutt, A. M., Armendariz, V., Arias, F., Diaz Santos, A. L., Zink, D., Vuong, K. D., & Rossetti, M. A. (2023). Culturally and linguistically informed neuropsychological evaluation protocol for primarily Spanish-speaking adults. *Archives of Clinical Neuropsychology*, *38*(3), 408–432. [CrossRef] [PubMed]

Substance Abuse and Mental Health Services Administration (SAMHSA). (2014). *SAMHSA's concept of trauma and guidance for a trauma-informed approach* (HHS Publication No. [SMA] 14-4884). Available online: https://www.health.ny.gov/health_care/medicaid/program/medicaid_health_homes/docs/samhsa_trauma_concept_paper.pdf (accessed on 14 September 2026).

van den Broek, M. (2025). Neuropsychology. In N. Agrawal, & N. Poole (Eds.), *Essential neuroscience for psychiatrists* (pp. 174–197). Cambridge University Press. [CrossRef]

van der Linde, I., & Bright, P. (2018). A genetic algorithm to find optimal reading test word subsets for estimating full-scale IQ. *PLoS ONE*, *13*(10), e0205754. [CrossRef] [PubMed]

Veiel, H. O. F., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, *13*(3), 356–368. [CrossRef] [PubMed]

Vonk, J. M. J., Gross, A. L., Zammit, A. R., Bertola, L., Avila, J. F., Jutten, R. J., Gaynor, L. S., Suemoto, C. K., Kobayashi, L. C., O'Connell, M. E., Elugbadebo, O., Amofa, P. A., Staffaroni, A. M., Arce Rentería, M., Turney, I. C., Jones, R. N., Manly, J. J., Lee, J., & Zahodne, L. B. (2022). Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). *PLoS ONE*, *17*(2), e0264166. [CrossRef] [PubMed]

Watt, S., Ong, B., & Crowe, S. F. (2018). Developing a regression equation for predicting premorbid functioning in an Australian sample using the national adult reading test. *Australian Journal of Psychology*, *70*(2), 186–195. [CrossRef]

Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale—Revised*. Psychological Corporation.

Weuve, J., Proust-Lima, C., Power, M. C., Gross, A. L., Hofer, S. M., Thiébaut, R., Chêne, G., Glymour, M. M., Dufouil, C., & MELODEM Initiative. (2015). Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimer's & Dementia*, *11*(9), 1098–1109. [CrossRef] [PubMed]

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18. [CrossRef] [PubMed]

Wilkinson, G. S., & Robertson, G. J. (2017). *WRAT-5: Wide range achievement test professional manual*. Psychological Assessment Resources.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, *46*(6), 1554–1555. [CrossRef]

Wolfers, T., Doan, N. T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., Buitelaar, J. K., Ueland, T., Melle, I., Franke, B., Andreassen, O. A., Beckmann, C. F., Westlye, L. T., & Marquand, A. F. (2018). Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry*, *75*(11), 1146–1155. [CrossRef] [PubMed]

Wu, Y., Jia, M., Xiang, C., Lin, S., Jiang, Z., & Fang, Y. (2022). Predicting the long-term cognitive trajectories using machine learning approaches: A Chinese nationwide longitudinal database. *Psychiatry Research*, *310*, 114434. [CrossRef] [PubMed]

Yi, D., Seo, E. H., Han, J. Y., Sohn, B. K., Byun, M. S., Lee, J. H., Choe, Y. M., Ahn, S., Woo, J. I., Jun, J., & Lee, D. Y. (2017). Development of the Korean adult reading test (KART) to estimate premorbid intelligence in dementia patients. *PLoS ONE*, *12*(7), e0181523. [CrossRef] [PubMed]

Yuan, C., Linn, K. A., & Hubbard, R. A. (2023). Algorithmic fairness of machine learning models for Alzheimer disease progression. *JAMA Network Open*, *6*(11), e2342203. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.