





Research Article

Wechsler Adult Intelligence Scale – IV México versus U.S. versions in the assessment of Mexican Americans

Gabriela Ontiveros  and Philip Gerard Gasquoiné 

Department of Psychological Science, University of Texas Rio Grande Valley, Edinburg, TX, USA

Abstract

Objective: To delineate score differences between the Wechsler Adult Intelligence Scale-IV (WAIS-IV) and the WAIS-IV México in the assessment of balanced bilingual Mexican Americans and to determine the efficacy of five hold measures in predicting summary scores in each version. **Methods:** Hold measures were WAIS-IV Information, Vocabulary, and Matrix Reasoning subtests, picture vocabulary, and the Test of Premorbid Function (English)/Word Accentuation Test (Spanish). Using a repeated measures design, 60 neurologically intact participants were tested in a counterbalanced order, with WAIS-IV version as the repeated measure (mean intertest interval = 5.68 days). To minimize practice effects, the five visual-perceptual subtests, which contain the same items in each version, were administered only once during the initial session. **Results:** All mean WAIS-IV México index/subtest scores were significantly higher than the U.S. equivalents (Full-Scale IQ by about .5 SD). Unexpectedly, most (83%) participants educated in the US to at least a high school level had numerically equal or higher scores on the U.S. version. Means on WAIS-IV language format indices/subtests were lower than those of visual-perceptual format indices/subtests within both versions (excepting Processing Speed Index/subtests in the U.S. version). All hold measures significantly predicted WAIS-IV summary scores for the U.S. version. Similarly for the México version, except for the Word Accentuation Test. **Conclusions:** When evaluating a balanced bilingual Mexican American, opting for the WAIS-IV México version will yield higher scores across the Full-Scale IQ, indices, and all core subtests unless the patient was educated in the US to at least a high school level.

Keywords: Spanish speakers; Spanish/English bilingualism; national norms; hold measures; preexisting neuropsychological skill estimate; language

(Received 12 May 2024; final revision 25 July 2024; accepted 3 September 2024; First Published online 17 October 2024)

Introduction

The Wechsler intelligence scales are the most frequently used measure of intelligence/achievement among U.S./Canadian clinical neuropsychologists, whether an evaluation is conducted in English (Rabin et al., 2016) or Spanish (Gasquoiné et al., 2021). While Spanish language equivalents of the child scales have been adapted and normed in the continental US since the fourth edition (Wechsler, 2005), continental U.S. Spanish language equivalents of the adult scales have never been developed. Instead, U.S. clinical neuropsychologists who evaluate patients in Spanish have been forced to rely on Wechsler Adult Intelligence Scales (WAIS) adapted and normed in various Spanish-speaking regions/countries.

The first of these was the Escala de Inteligencia Wechsler para Adultos (EIWA; Wechsler, 1968) adapted and normed in Puerto Rico. When the EIWA was utilized to evaluate Spanish-speaking Hispanic Americans residing in the continental US, it generated scores more than 20 points higher than English language WAIS scores (Melendez, 1994). Subsequent comparisons of various nationalized Spanish language translations/adaptions of various WAIS editions versus the U.S. version using hypothetical raw

scores (Funes et al., 2016) or those from 305 highly educated, Columbian, corporate executives (Duggan et al., 2019) also found higher summary scores (indices; Full-Scale IQ) in the Spanish language versions by varying amounts. These score differences have been attributed to lower national mean educational levels in the Spanish-speaking countries than that within the US (Funes et al., 2016).

Current Spanish language adaptations/translations of the WAIS-IV (Wechsler, 2008) include those for Spain (Wechsler, 2012), Chile (Wechsler, 2013), México (Wechsler, 2014), and Colombia (Wechsler, 2016). Aside from education, another important national characteristic that could potentially account for the higher WAIS summary scores is that residents of these countries are primarily monolingual Spanish speakers. In contrast, most (63%) persons who make up the census category of Hispanic Americans (Latine), defined as U.S. residents who trace ancestry to any Spanish-speaking country, are bilingual to some degree with only 5% being monolingual Spanish speakers (Duffin, 2022).

Hispanic Americans became a federally recognized census category in 1977 and were subsequently included in U.S. Wechsler

Corresponding author: Gabriela Ontiveros; Email: gabriela.ontiveros01@utrgv.edu

Cite this article: Ontiveros G. and Gasquoiné 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: 819–827, <https://doi.org/10.1017/S135561772400050X>

© The Author(s), 2024. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

intelligence scale normative studies. As a group, they had mean scores about 0.5 *SD* below non-Hispanic European Americans on language format measures (verbal IQ), with no difference on visual-perceptual format measures (performance IQ; Neisser et al., 1996; Puente & Salazar, 1998). Visual-perceptual format measures are those that primarily require the processing of geometric form stimuli. The language versus visual-perceptual format distinction mirrors the lateralization of the dominant versus non-dominant cerebral hemispheres, respectively, in the human brain.

Subsequent research has confirmed a bilingualism effect on language, but not visual-perceptual, formatted test scores in neurologically intact bilingual Hispanic Americans across multiple neurocognitive measures (Gasquoine et al., 2010; Gasquoine, 2016). As illustration, language proficiency in either language was shown to positively and significantly correlate with language format neuropsychological measures of executive function (digits forwards [but not backwards]; Stroop interference; letter fluency) and delayed memory (Story Memory; California Verbal Learning Test). In contrast, language proficiency in either language did not significantly correlate with neuropsychological visual-perceptual format executive function (spatial span forwards and backwards; Wisconsin Card Sorting Test categories and errors) or delayed memory (visual memory) measures (Ontiveros & Gasquoine, 2023).

Theoretical explanations for this language effect include the frequency-lag hypothesis, whereby bilinguals lag behind monolinguals in word usage frequency, particularly in their non-dominant language (Gollan et al., 2011). Another theory posits that bilinguals activate both languages concurrently during language processing (Green, 1998), leading to a larger search base for unfamiliar word retrieval compared to monolinguals, thereby prolonging the process, and increasing the chances of error.

Clinical neuropsychological assessments should always be conducted in Spanish with Spanish-dominant Hispanic Americans to avoid test bias (i.e., scores reflecting language proficiency rather than intelligence: American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014). Spanish versus English language assessments are optional when the patient is a balanced bilingual. The current study had a primary goal of delineating score differences between these two options by comparing mean group differences between the WAIS-IV México and U.S. versions for a sample of balanced bilingual Mexican Americans.

Mexican Americans are the largest (62%) and most well-established of the Hispanic American national groupings (U.S. Department of Health and Human Services, 2020). Around the time the WAIS-IV México was being normed, the World Bank (2010) estimated that the Mexican population had an average education attainment of 9 years, with 16% of adults completing post-secondary education. In contrast, U.S. adults averaged 13 years of school with 42% completing post-secondary education. Given that the average IQ in both versions is 100, it was expected that this would be attained by individuals with fewer years of education on the México than on the U.S. version. For any given individual, WAIS summary scores would be expected to trend higher for the México version and balanced bilinguals should perform better on visual-perceptual format summary measures (i.e., Perceptual Reasoning [PRI] and Processing Speed [PSI] Indices) than those with a language format (i.e., Verbal Comprehension [VCI] and Working Memory [WMI]) for both versions.

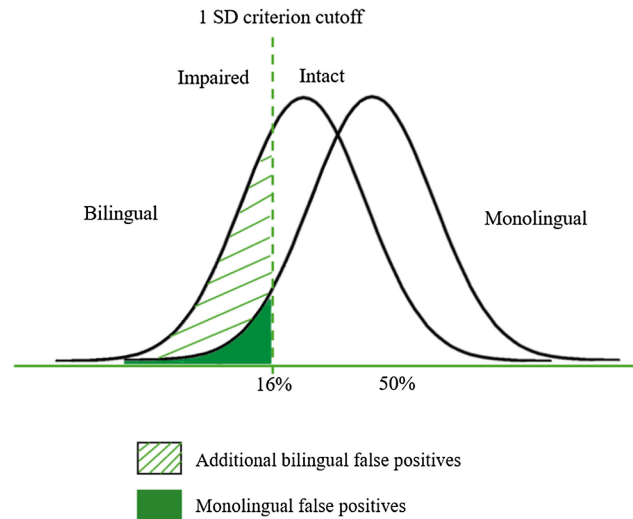


Figure 1. Hypothetical distributions of language-formatted neuropsychological test scores for neurologically intact bi- versus monolinguals.

Estimating preexisting neuropsychological skill level

The purpose of many clinical neuropsychological assessments is to identify acquired neurocognitive impairment, conceptualized as a decline by a certain amount (e.g., 1 *SD*) from a preexisting neuropsychological skill level. Typically, this preexisting neuropsychological skill level has to be estimated as the assessment takes place post-injury. A common practice is to place the preexisting estimate at the 50th percentile (Heaton et al., 2004), thereby effectively changing the definition of acquired neurocognitive impairment to that of a statistically low score. Figure 1 shows how this approach can hypothetically increase the number of false positives (i.e., neurologically intact individuals misdiagnosed with brain injury) among U.S. ethnic/linguistic minorities, whose neurologically intact test score distributions fall below that of a normative monolingual, European American grouping.

One solution is the creation of separate normative tables for U.S. ethnic/linguistic minority groupings, but these perpetuate minority group stereotypes and are impractical due to high cost and difficulties in scientifically defining homogeneous groupings (Gasquoine, 2009, 2022). A more practical solution is to estimate the preexisting neuropsychological skill level from post-injury neuropsychological test scores that are known to be relatively resistant to the type of brain injury suspected, called *hold* measures (Gasquoine & Gonzalez, 2012). These hold measures are likely equally subject to whatever cross-cultural variables (e.g., educational level; bilingualism) account for the lower neuropsychological test scores in the ethnic/linguistic minority grouping. The concept of hold measures as resistant to mental deterioration predates the popularization of clinical neuropsychology, being initially utilized as vocabulary scores in schizophrenia (Babcock, 1930) and, after the first edition of the WAIS was published, with hold versus “don’t hold” subtests in aging (Levi et al., 1945).

Wechsler hold core subtests are Information, Vocabulary, and Matrix Reasoning (Donders et al., 2001), with another popular hold measure for non-aphasic brain injury involving the pronunciation of phonetically irregular words. Research involving determination of the best hold measure to use with various brain

injury groupings has been hampered by the lack of “true” measures of preexisting neuropsychological skill level. Another research methodology to determine the efficacy of various hold measures is to compare their correlations with Wechsler summary scores, especially Full-Scale IQ, in neurologically intact samples (Bright & van der Linde, 2020). In the current study, the efficacy of hold word pronunciation and Wechsler core subtest scores was evaluated in terms of their correlation with Wechsler summary scores in both WAIS-IV México and U.S. versions. It was expected that optimal hold scores would differ for language versus visual-perceptual format WAIS-IV summary scores according to format match.

Methods

Participants

Study participants comprised consecutive (i.e., no exclusions), community dwelling, residents of the Rio Grande Valley in the Texas international borderlands region who met the following inclusion criteria: (a) ≥ 18 years of age, (b) of Mexican ancestry, (c) bilingual in Spanish and English as initially demonstrated by the ability to converse in both languages, and (d) the self-reported absence of a neurological or psychiatric disorder for which they had been hospitalized or were taking psychoactive medications. Recruitment occurred through informal channels, relying on individuals within the community to share Spanish and English language flyers. All participants gave informed consent, and study procedures were approved by the Institutional Review Board of the University of Texas Rio Grande Valley. The research was completed in accordance with the Helsinki Declaration.

Participant characteristics ($N = 60$) are summarized in Table 1. There were 36 males and 24 females ranging in age from 18 to 63 years ($M = 43.87$; $SD = 14.41$). Most (42) were born in México, subsequently living in the US for 12 to 48 years ($M = 31.43$; $SD = 10.74$). Participant education ranged from 3 to 18 years ($M = 10.18$; $SD = 3.64$) with 24 being educated in the US, 34 in México, and 2 in both countries. Those educated in the US had more years of education (range = 11–18; $M = 13.04$; $SD = 2.10$) than those educated in México (range = 3–12; $M = 8.06$; $SD = 3.14$). Household incomes ranged from \$10,000 to \$120,000 ($M = \$31,033$; $SD = \$16,511$). The median of \$30,000 was well below the median for the region at \$48,000 (U.S. Census Bureau, 2023).

Current preferred language for conversation was Spanish for 33 participants, English for 11, and either for 16. Spanish was the first language for every participant, and for 46, it was the predominate language spoken at home. Age of second language acquisition ranged from 4 to 36 years ($M = 13.2$; $SD = 8.41$).

Measures

Language proficiency and dominance

The Woodcock-Muñoz Language Survey-Revised (WMLS-R; Woodcock et al., 2005) Picture Vocabulary subtest ($M = 100$; $SD = 15$) was used to objectively measure language proficiency in each language. The WMLS-R English version was normed within the US on 8,818 participants, selected according to U.S. census projections issued in 1996 for the year 2000. The Spanish version was normed on 1,157 Spanish speakers from various countries in Latin America and the US (7%). Spanish data was equated to the English norms using the Rasch modeling.

As in Gasquoine et al. (2007), the Spanish and English WMLS-R Picture Vocabulary standard scores were subtracted from each other to provide a measure of language dominance. A priori, it was

Table 1. Participant characteristics ($n = 60$)

Characteristics	n (%)	M	SD
Female	24 (40)		
Male	36 (60)		
Age (years)		43.87	14.41
Country of birth: México	42 (70)		
(Years in US ^a)		31.43	10.74
US	18 (30)		
Years of education		10.18	3.64
Country of education: México	34 (57)		
US	24 (40)		
Both	2 (3)		
Occupation: Professionals/managers	11 (18)		
Service/sales	18 (30)		
Agricultural worker	13 (22)		
Trades worker	16 (27)		
Student	2 (3)		
Household income		\$31,033	\$16,511
First language: Spanish	60 (100)		
Age (years) second language acquired		13.20	8.41
Language predominately spoken at home:			
Spanish	46 (77)		
English	6 (10)		
Both	8 (13)		
Current preferred language: Spanish	33 (55)		
English	11 (18)		
Both	16 (27)		

^aFor participants born in México.

decided that if any participant had a dominance score $\geq \pm 15$ (i.e., $\pm 1 SD$), they were to be classified as language dominant and excluded from post hoc data analysis. No participant was so excluded.

WAIS-IV and WAIS-IV México

The WAIS-IV has 10 core subtests ($M = 10$; $SD = 3$) that generate five summary scores ($M = 100$; $SD = 15$), namely four indices (VCI, PRI, WMI, and PSI) and the Full-Scale IQ (FSIQ). The scale was normed in the US on a sample of 2,200 individuals selected to be representative of the 2005 census in terms of age (13 age-bands), gender, race/ethnicity (White, African American, Hispanic, Asian, or “Other”), education attainment of self (for ages 20–90) or parents (for ages 16–19), and geographic region.

The first iteration of the Mexican version of the WAIS (Wechsler, 2003) was a direct translation of the U.S. third edition with minor changes to instruction and item hierarchy. It was normed in México ($N = 970$) on an ill-defined sample and users were given the option of selecting either the Mexican or U.S. norms. The Mexican norms produced full-scale IQs on average 12 points higher (Suen & Greenspan, 2009).

The WAIS-IV México was normed on a sample of 1,450 Mexican residents (52% female), ages 16–90 years (13 age-bands) from seven states. These seven states were chosen because universities in each coordinated the standardization (Table 2). The technical manual provided no information on the educational or racial/ethnic background of the standardization sample.

The México version format is exactly the same as the US except for ordering (reflecting relative difficulty) item changes in Información (Information), Semejanzas (Similarities), Vocabulario (Vocabulary), Aritmética (Arithmetic), Matrices (Matrix Reasoning), and Rompecabezas Visual (Visual Puzzles) subtests. Substantive item changes were in Información and Vocabulario that had both similar and substituted items. For example in Información, item 23, “Who created the character

Table 2. Zones, states, and universities of Wechsler Adult Intelligence Scale-IV México standardization sample

Zone	State (University)	Participants
Centro (Central)	Distrito Federal (Universidad Nacional Autónoma de México)	297
	Hidalgo (Universidad Autónoma del Estado de Hidalgo)	199
	Guanajuato (Universidad de Guanajuato)	182
	Puebla (Universidad de Anáhuac de Puebla)	171
Norte (North)	Sinaloa (Universidad Autónoma de Sinaloa)	299
	Nuevo León (Universidad de Monterrey)	214
Sur (South)	Yucatán (Universidad Autónoma de Yucatán)	88
Total		1450

Sherlock Holmes?” and “¿Quién creó el personaje Sherlock Holmes?” are the same. Conversely, “Who was the president of the United States during the Civil War?” (item 11) differs from its substitute “¿Quién fue Emiliano Zapata?” (item 7). In Vocabulario, item 19, “generate,” and item 17, “generar,” match but “tirade” (item 28) differs from its substitution “osado” (item 27).

Other item changes across the other subtests were minor. In Aritmética, names of individuals were non-anglicized, and in item 18, “pies” was replaced by “dulces” (candy). For Retención de dígitos (Digit Span), the numbers were the same, but Spanish has five digits between 1 and 9 that have two syllables (i.e., cuatro, cinco, siete, ocho, and nueve), whereas English has only one (i.e., seven). Average digit span forwards is thus typically shorter in Spanish than English (e.g., 6.4 vs. 7.2; Naveh-Benjamin & Ayres, 1986).

Word pronunciation

The Test of Premorbid Functioning (TOPF; The Psychological Corporation, 2009) was co-normed ($M = 100$; $SD = 15$) with the WAIS-IV and requires the participant to correctly read 70 phonologically irregular (e.g., “mosquito”; “paradigm”) English words. The participant raw score was the number of correctly pronounced words.

In Spanish, grapheme to phoneme mapping is regular, so individuals with little lexical knowledge can pronounce most words. Spanish pronunciation becomes less regular in terms of accentuation, and when rules of accentuation deviate, an acute accent is placed above the syllable to cue the reader. Using this approach, the Test de Acentuación de Palabras (Word Accentuation Test [WAT]: Del Ser et al., 1997) was developed that requires participants to correctly pronounce 30 infrequent Spanish words written without accentuation marks (e.g., “bulgaro”; “abogacia”). The only available age and education corrected norms ($M = 10$; $SD = 3$) for the WAT come from 700 neurologically intact residents of Spain ≥ 18 years of age (Del Pino et al., 2018). A 40-item version containing many of the original words was adapted for use in the US (Krueger et al., 2006), but no normative data was available, so it was not used here. The participant raw score was the number of correctly pronounced (either European or Latin American Spanish) words.

Emotional state

Psychometric measures of emotional state were provided by the Beck Depression Inventory-II (BDI-II; Beck et al., 1996) and the Beck Anxiety Inventory (BAI; Beck & Steer, 1993) that are available in both Spanish and English. A priori, it was decided that any participant with a BDI-II score > 13 (classified as $> minimal$) or

BAI score > 15 ($> mild$) in either language was to be excluded post hoc from the data analysis. In Spanish, BDI-II scores ranged from 0 to 10 ($M = 3.73$; $SD = 2.26$) and BAI scores from 0 to 6 ($M = 1.93$; $SD = 1.72$). In English, BDI-II scores ranged from 0 to 11 ($M = 3.78$; $SD = 2.33$) and BAI scores ranged from 0 to 7 ($M = 2.02$; $SD = 1.74$). No participants were excluded.

Procedure

Participants attended two sessions conducted in Spanish or English in a counterbalanced design, to which they were randomly assigned. During each session, the first author, who is of Mexican ancestry, bilingual, and a local resident, and the participant exclusively communicated in the assigned language. To maintain language consistency, any cross-language intrusions were gently corrected by reminding the participant to respond only in the language assigned.

The WAIS-IV subtests that make up the PRI and PSI have the same items in each version and so were administered only once at the first session. This helped to minimize practice effects that trend higher for these visual-perceptual format subtests (average gain of 1.16 points) than for the language format subtests of VCI and WMI (0.75 points; Estevis et al., 2012). The single set of raw scores from these subtests were converted to scale scores using the norms for each WAIS-IV version.

The first session in Spanish or English consisted of the following: (a) consent form, (b) demographic questionnaire, (c) WMLS-R Picture Vocabulary subtest, (d) WAIS-IV 10 core subtests, (e) word pronunciation test, (f) BDI-II, and (g) BAI. Intertest intervals ranged from 3 to 16 days ($M = 5.68$; $SD = 3.18$). The second session in the other language consisted of the following: (a) WMLS-R Picture Vocabulary subtest, (b) WAIS-IV five core subtests comprising VCI and WMI, (c) word pronunciation test, (d) BDI-II, and (e) BAI. Upon completion of the second session, each participant received the study incentive, a \$49.00 grocery gift card.

Analytic strategy

All data analysis was performed using SPSS version 29 software (IBM Corp., 2023). There was no missing data, and statistical significance was established at the 0.05 α level. Two repeated measures Multivariate Analysis of Variance (RM-MANOVA) with language of administration as the repeated measure were analyzed. The first was a 2×4 RM-MANOVA with the four WAIS-IV index scores and the second with the 10 WAIS-IV core subtests. In the case of a significant main effect for WAIS-IV index scores or core subtests, Bonferroni post hoc tests were conducted to determine where the significant differences lay.

Two correlational matrices, one in English and one in Spanish, compared the four WAIS-IV indices, the full-scale IQ, and the five hold measures of preexisting neuropsychological skill level. Larger, positive Pearson Product Moment correlations are indicative of optimal predictors of WAIS-IV summary scores.

Results

Language proficiency and dominance

Mean WMLS-R Picture Vocabulary standard scores in Spanish ranged from 69 to 100 ($M = 85.87$; $SD = 7.37$) and 68 to 105 ($M = 86.05$; $SD = 8.99$) in English. These mean WMLS-R scores were about 1 SD below the national U.S. respective monolingual

means of 100. The sample was well balanced with a dominance score (México minus U.S. mean) of $-.18$.

WAIS-IV México versus U.S. versions

Table 3 presents the range, mean, and standard deviation for the five WAIS-IV summary scores for the México and U.S. versions. All mean scores for the México version were higher than their U.S. equivalent, with differences ranging from 4.12 for VCI to 10.25 for PSI (Figure 2). The mean FSIQ for the sample was 7.40 points higher (i.e., nearly $.5$ SD) for the México (89.80) than the U.S. (82.40) version.

In a 2×4 RM-MANOVA with WAIS-IV version as the repeated measure and index scores as the within subject variable, the main effect of version found the mean (the average of the four index scores not the FSIQ) for the México version of 91.39 to be significantly higher than that of the U.S. version mean at 85.24, $F(1, 59) = 85.15$, $p < .001$, $\eta^2 = .59$. The main effect of WAIS-IV index scores was also significant, $F(3, 57) = 4.92$, $p = .004$, $\eta^2 = .21$. Bonferroni post hoc comparisons showed the PRI mean of 89.97 was significantly higher than the VCI mean of 86.99 ($p = .02$). The interaction was also significant, $F(3, 57) = 13.49$, $p < .001$, $\eta^2 = .42$. Paired samples t tests illustrated that all mean summary scores for the México version were significantly higher than their U.S. equivalents at $p < .001$.

Table 4 shows the range, mean, and standard deviation for the 10 WAIS-IV core subtest scaled scores in the México and U.S. versions. All mean scores for the México version were higher than their U.S. equivalents, with differences ranging from 0.48 for Block Design to 2.00 for Coding (Figure 3). For an RM-MANOVA, the main effect of version found the México mean of 8.45 was significantly higher than the U.S. mean of 7.41, $F(1, 59) = 89.50$, $p < .001$, $\eta^2 = .60$. The main effect of WAIS-IV subtest scaled scores was also significant, $F(9, 51) = 4.28$, $p < .001$, $\eta^2 = .43$. Bonferroni post hoc comparisons showed the mean for Block Design of 8.41 was significantly higher than the means for Vocabulary of 7.55 ($p = .03$) and Information of 7.69 ($p = .03$). The interaction was also significant, $F(9, 51) = 7.52$, $p < .001$, $\eta^2 = .57$. Paired samples t tests illustrated that all mean subtest scores for the México version were significantly higher than their U.S. equivalents ($p < .002$).

On the México version, all language format subtest scores were lower than the visual-perceptual format subtest scores. Similarly for the U.S. version, excepting Symbol Search and Coding scores that make up the PSI.

Individual analysis

Although most participants scored higher on the WAIS-IV México, 9 of 60 (15%) participants scored higher on the U.S. version and 3 scored equally on both language versions. All 12 of these participants had at minimum 12 years of education in the US (one was educated in both countries). In contrast the 8 participants who completed 12 years of education in México, all had higher FSIQ scores on the México version.

A U.S. over México version score superiority occurred most frequently on the VCI (17/60 = 28%). All 17 of these participants had at minimum 12 years of education in the US (one was educated in both countries). There were six other participants with a U.S. high school or better education whose VCI scores were higher on the México version. There were nine (15%) participants with higher U.S. version WMI scores (all also had higher VCI scores). Three participants (5%) scored higher on the U.S. version PRI and two (3%) on the PSI.

Table 3. Wechsler Adult Intelligence Scale-IV México and U.S. version summary scores

Summary scores	México			U.S.			t test
	Range	M	SD	Range	M	SD	
VCI	65–112	89.05	7.37	58–122	84.93	13.56	3.52*
PRI	73–128	92.35	11.48	60–127	87.58	13.06	12.22*
WMI	62–112	90.10	10.55	58–111	84.63	13.13	7.49*
PSI	75–116	94.07	9.45	56–117	83.82	15.29	11.25*
FSIQ	69–118	89.80	9.25	53–119	82.40	14.46	8.49*

Note: VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full-Scale IQ.

* $p < .001$.

Hold measures

Table 5 shows the Pearson Product Moment correlations of five hold measures with WAIS-IV México summary scores. The WAT was negatively correlated with all WAIS-IV summary scores. All participants had difficulty with this test such that the highest raw score was only 11 of 30 ($M = 5.55$; $SD = 2.45$). All other hold measures were significant predictors of FSIQ, ranging from .70 (WMLS-R Picture Vocabulary) to .85 (WAIS-IV Matrix Reasoning). For the language format indices, the best predictor of VCI was WAIS-IV Vocabulary (.84), and for WMI, it was WMLS-R Picture Vocabulary (.69). For the visual-perceptual format indices, the best predictor of PRI was WAIS-IV Matrix Reasoning (.93), and for PSI, it was WAIS-IV Vocabulary (.63).

Table 6 shows the Pearson Product Moment correlations of five hold measures with U.S. WAIS-IV summary scores. All hold measures were significant predictors of FSIQ, ranging from .83 (WAIS-IV Vocabulary) to .89 (WAIS-IV Information). For the language format indices, WAIS-IV Information (.94) proved to be the best predictor of VCI, and for WMI, three (WMLS-R Picture Vocabulary, WAIS-IV Information, and Matrix Reasoning) were at .77. For the visual-perceptual format indices, the best predictor of PRI was WAIS-IV Matrix Reasoning (.96), and for PSI, it was TOPF (.78).

Discussion

The WAIS-IV México generated a mean FSIQ about $.5$ SD higher than that for the U.S. version. Higher scores on the México version were pervasive across all indices and core subtests, suggesting the effect is likely due to the national difference in mean years of education (9 in México vs. 13 in the US). Bucking the trend, individual qualitative analysis showed most participants educated at least to high school in the US had equal or higher FSIQs on the U.S. version. This effect was primarily driven by the VCI score on which 19 of the 23 participants (83%) with a minimum U.S. high school education had higher or equal scores on the U.S. version. VCI includes the two subtests (Information and Vocabulary) where substantive cultural changes were made in item content to the México version.

Previous research on U.S. normed Wechsler intelligence scales has shown Hispanic Americans score lower on language than visual-perceptual format measures (Gasquoin et al., 2010; Neisser et al., 1996; Puente & Salazar, 1998). For balanced bilingual Mexican Americans, language proficiency had been previously shown to correlate with multiple language format neuropsychological measures of memory and executive function but not with any such visual-perceptual format measure (Ontiveros &

Table 4. México and U.S. version Wechsler Adult Intelligence Scale-IV subtests scaled scores

Subtests	México			U.S.			t-test
	Range	M	SD	Range	M	SD	
Information	3–12	8.10	1.80	3–15	7.28	2.68	3.45*
Vocabulary	3–11	7.93	1.85	3–12	7.17	2.27	4.69**
Similarities	4–15	8.27	1.97	3–15	7.55	2.79	2.96**
Digit Span	4–13	8.27	2.05	2–13	7.27	2.31	6.65*
Arithmetic	3–13	7.95	2.30	2–14	7.32	2.61	4.32*
Block Design	5–18	8.65	2.38	5–16	8.17	2.16	5.54*
Matrix Reasoning	4–16	8.85	2.37	2–17	7.80	2.99	8.44*
Visual Puzzles	5–14	8.60	1.94	3–14	7.68	2.14	12.02*
Symbol Search	6–14	9.18	2.07	3–15	7.23	3.15	11.12*
Coding	4–13	8.67	1.84	1–12	6.67	2.72	13.14*

*p < .001; **p = .002.

Gasquoine, 2023). In concert, the VCI mean combined across both versions was significantly lower than the combined PRI mean. This was the only significant difference among the combined index scores, but the trend was evident across the WMI and PSI indices that have greater executive function load, excepting that the U.S. version PSI was the lowest index score. U.S. version Coding had the lowest and Symbol Search the third lowest (above Vocabulary) subtest score. Within the repeated measures design, PSI subtests were only administered once, so any score difference between the two WAIS-IV versions is wholly attributable to differences in relative scaled scores from the normative samples.

Lower scores for language versus visual-perceptual format indices/subtests advocate for the selection of two differently

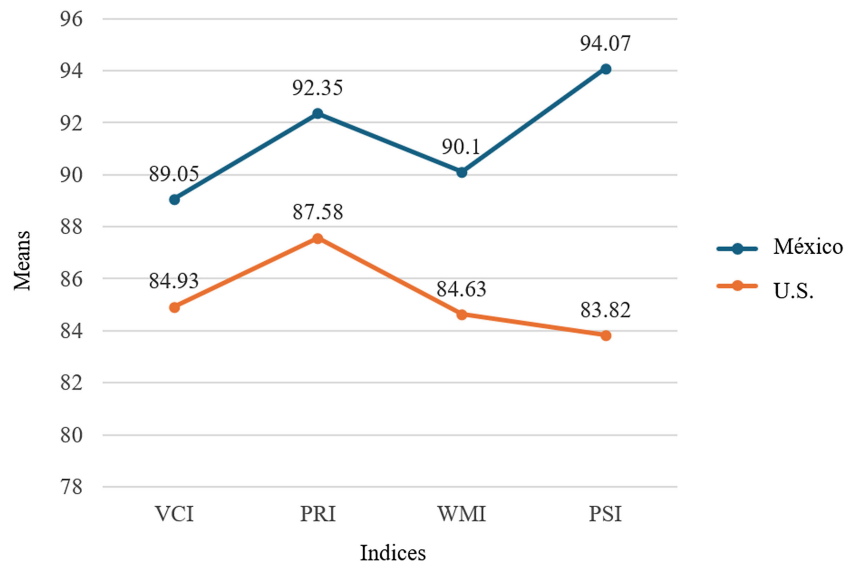


Figure 2. México and U.S. version Wechsler Adult Intelligence Scale-IV index scores. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

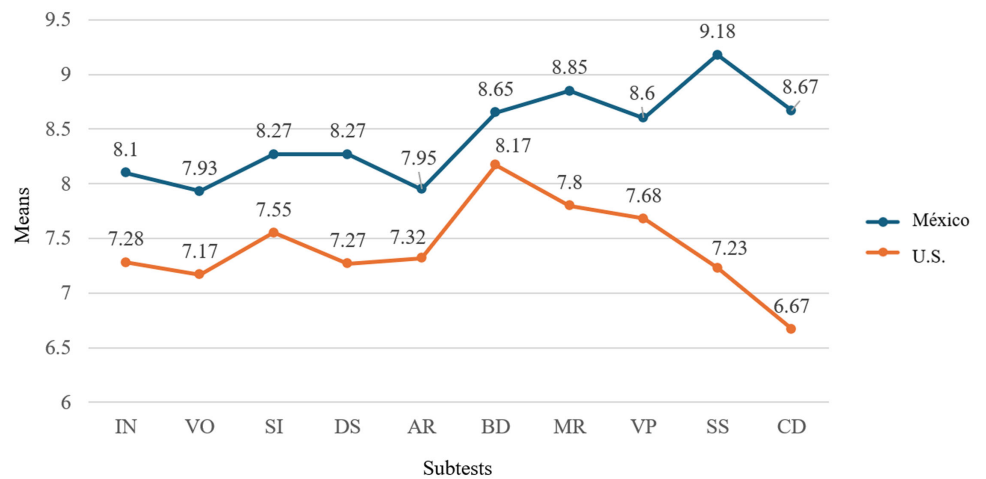


Figure 3. México and U.S. version Wechsler Adult Intelligence Scale-IV mean subtest scores. IN = Information, VO = Vocabulary, SI = Similarities, DS = Digit Span, AR = Arithmetic, BD = Block Design, MR = Matrix Reasoning, VP = Visual Puzzles, SS = Symbol Search, CD = Coding.

Table 5. Pearson product moment correlations of hold measures with Wechsler Adult Intelligence Scale-IV summary scores for México version

Hold measures	VCI	PRI	WMI	PSI	FSIQ
WMLS-R PV	.68*	.58*	.69*	.59*	.70*
WAIS-IV VO	.86*	.57*	.62*	.63*	.73*
WAIS-IV IN	.84*	.68*	.65*	.59*	.78*
WAIS-IV MR	.68*	.93*	.68*	.59*	.85*
WAT	-.63*	-.56*	-.64*	-.60*	-.67*

Note: WMLS-R PV = Woodcock Muñoz Language Survey-Revised Picture Vocabulary in Spanish, VO = Vocabulary subtest, IN = Information subtest, MR = Matrix Reasoning subtest, WAT = Word Accentuation Test, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full-Scale IQ.

* $p < .001$.

Table 6. Pearson product moment correlations of hold measures with WAIS-IV summary scores for U.S. version

Hold measures	VCI	PRI	WMI	PSI	FSIQ
WMLS-R PV	.85*	.74*	.77*	.74*	.86*
WAIS-IV VO	.93*	.66*	.68*	.70*	.83*
WAIS-IV IN	.94*	.78*	.77*	.69*	.89*
WAIS-IV MR	.76*	.96*	.77*	.65*	.88*
TOPF	.84*	.74*	.73*	.78*	.86*

Note: WMLS-R PV = Woodcock Muñoz Language Survey-Revised Picture Vocabulary in English, WAIS-IV VO = Vocabulary subtest, WAIS-IV IN = Information subtest, WAIS-IV MR = Matrix Reasoning subtest, TOPF = Test of Premorbid Functioning, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, FSIQ = Full-Scale IQ.

* $p < .001$.

formatted hold measures to estimate language versus visual-perceptual preexisting skill level in balanced bilingual Mexican Americans. There was limited support for this, as all hold measures (except the WAT) were significantly positively correlated with all summary scores, especially in the U.S. version. Nevertheless, in both versions, the sole visual-perceptual hold measure (Matrix Reasoning) had the highest correlation with PRI (of which it is part) and the lowest (tied with WMLS-R Picture Vocabulary in the México version) with VCI.

The missing accentuation word pronunciation approach as measured by the WAT proved ineffective as a positive predictor of WAIS-IV México summary scores. There was range restriction as indicated by the highest raw score reaching only 11 out of 30 items. European Spanish as spoken in Spain exhibits notable differences from Spanish spoken in Latin America (Stewart, 2012). Disparities encompass pronunciation, vocabulary, grammar, and cultural nuances. In terms of pronunciation, variations exist in the treatment of certain consonants and vowels, such as the “lisping s” sound for “c” and “z” in Spain, compared to the standard “s” sound in Latin America. Vocabulary discrepancies are also apparent, with different words or meanings prevailing in each variant. For instance, “coche” for “car” in Spain contrasts with “carro” in México. Some of the WAT words from Spain had no cultural relevance in the Rio Grande Valley (e.g., grisú [firedamp]; pifano [fife]). Grammatical rules largely align but may differ in usage, like the preference for “vosotros” or “ustedes” for the plural “you” in Spain and México, respectively. Spanish speakers in the US come from multiple countries and consequently use a variety of different linguistic conventions in spoken and written language. A goal of future test development is to adapt a missing accentuation

word pronunciation test suitable for assessing Hispanic Americans.

Study limitations and strengths

Another potential reason for the unsuitability of the WAT for this sample relates to the low educational level and socioeconomic status in comparison to other U.S. residents. Mean FSIQ scores for the WAIS-IV México (90:25th percentile) and U.S. versions (82:12th percentile) were likely driven by these demographics along with bilingualism (i.e., frequency-lag and/or simultaneous language activation) effects. The mean number of years of education for the sample (10) was higher than the national México mean (9), but whether the latter is representative of the WAIS-IV México normative sample is unknown. It is assumed that study results are applicable to other balanced bilingual Mexican Americans who reside in other parts of the country, but this cannot be determined from this study. Similarly, the demographics (e.g., age, income, education, acculturation) of balanced bilinguals as a specific grouping is not collected in census data, so it is unclear if the relatively low number of years of education and mean family income and high acculturation levels are representative of this population.

The critical feature of sample selection in this research design was to have bilingual balance as indicated by a dominance score close to zero. This was satisfied as evidenced by the WMLS-R Picture Vocabulary dominance score (Spanish minus English mean) of less than 1. The specification of dominance versus balance in a bilingual patient is better assessed as a comparison of language proficiency measures in each language rather than the more widely used self-report as the latter is dependent upon the reference frame of the patient (Tomoschuk et al., 2019).

Previous research has consistently shown that vocabulary scores for neurologically intact balanced bilingual individuals tend to be lower compared to those for monolinguals in both languages (Bialystok et al., 2010; Celik et al., 2022; Gasquoine, 2016). This trend was observed in the present sample, with mean WMLS-R Picture Vocabulary scores at 86 in each language, approximately 1 SD below the national monolingual mean of 100. Similarly, the mean WAIS-IV Vocabulary subtest score in the U.S. version was 7.17. Previous studies in this region of the country have found the same effect. As illustration, in a study comparing demographically matched balanced versus language-dominant groups of 3- to 7-year-old Mexican Americans, the mean scores on the WMLS-R Picture Vocabulary test were 88 in English and 89 in Spanish for the balanced bilingual group, compared to 105 in English and 104 in Spanish for the language-dominant groups (Weimer & Gasquoine, 2016). As these participants were barely entering the educational system, the effect is likely a characteristic of bilingualism rather than reflecting substandard educational practices.

Conclusions: clinical implications

When evaluating a balanced bilingual Mexican American patient, the clinician has a choice to make between national test versions. Administering both introduces the confound of practice effects. Opting for the México version typically yields significantly higher scores across the FSIQ (by about .5 SD), indices, and all core subtests. A notable exception is when a patient is educated in the US and has achieved at least a high school level in which case the U.S. version typically yields higher or comparable summary scores. Caution regarding the use of the WAIS-IV México pertains to the

characteristics of its normative sample whereby the educational level is unknown, and its regional representation was selective.

Except for diagnosing learning and intellectual developmental disorders (Fletcher & Miciak, 2024), clinical neuropsychologists generally find limited utility in the magnitude of summary scores like the FSIQ, preferring to analyze variations in scores among the indices or subtests in comparison to other neuropsychological measures. Therefore, a consideration for choice of version is the country in which the other neuropsychological tests in the battery were normed to facilitate cross-test comparisons.

Using performance-based estimates of preexisting neuropsychological skill level in place of the 50th percentile of published norms can reduce the number of false positives among groups of U.S. linguistic/ethnic minorities whose mean scores fall below the mean of monolingual, non-Hispanic European Americans (Gasquoine, 2009, 2022; Gasquoine & Gonzalez, 2012). For both WAIS-IV versions, this study showed that the hold measures of WMLS-R Picture Vocabulary, WAIS-IV Information, WAIS-IV Vocabulary, and WAIS-IV Matrix Reasoning were significant predictors of all summary scores. The hold measure assessing irregular word pronunciation (TOPF) was a significant predictor for the U.S. version, whereas the WAT did not for the México version. As balanced bilingual Mexican Americans tend to score lower on many language than visual-perceptual format measures, an optimal hold measure for estimating preexisting language skill level is one that is language formatted, whereas the optimal hold measure for estimating visual-perceptual skill level is the WAIS-IV Matrix Reasoning subtest.

Cross-nation comparisons of common neuropsychological tests shed light on an implicit assumption of the field, namely that norms are defined by country. As is apparent in the present study, there are significant differences in these normative databases that impact the determination of neurocognitive impairment. When evaluating immigrant groupings, the issue of whether to use norms from the originating or host country is an avenue that requires future research attention.

Funding statement. None.

Competing interests. None.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Babcock, H. (1930). An experiment in the measurement of mental deterioration. *Archives of Psychology*, 117, 1–105.
- Beck, A. T., & Steer, R. A. (1993). *Beck anxiety inventory manual*. Psychological Corporation.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II*. Psychological Corporation.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531.
- Bright, P., & van der Linde, I. (2020). Comparison of methods for estimating pre-morbid intelligence. *Neuropsychological Rehabilitation*, 30(1), 1–14.
- Celik, S., Kokje, E., Meyer, P., Frölich, L., & Teichmann, B. (2022). Does bilingualism influence neuropsychological test performance in older adults? A systematic review. *Applied Neuropsychology: Adult*, 29(4), 855–873.
- Del Pino, R., Peña, J., Ibarretxe-Bilbao, N., Schretlen, D. J., & Ojeda, N. (2018). Demographically calibrated norms for two pre-morbid intelligence measures: The word accentuation test and pseudo-words reading subtest. *Frontiers in Psychology*, 9, 1–11.
- Del Ser, T., González-Montalvo, J. I., Martínez-Espinosa, S., Delgado-Villalpalos, C., & Bermejo, F. (1997). Estimation of pre-morbid intelligence in Spanish people with the word accentuation test and its application to the diagnosis of dementia. *Brain and Cognition*, 33(3), 343–356.
- Donders, J., Tulskey, D. S., & Zhu, J. (2001). Criterion validity of new WAIS-III subtest scores after traumatic brain injury. *Journal of the International Neuropsychological Society*, 7(7), 892–898.
- Duffin, E. (2022). *Distribution of English language proficiency among Hispanics in the United States as of 2022*. <https://www.statista.com/statistics/639745/us-hispanic-english-proficiency/>.
- Duggan, E. C., Awakon, L. M., Loaiza, C. C., & Garcia-Barrera, M. A. (2019). Contributing towards a cultural neuropsychology assessment decision-making framework: Comparison of WAIS-IV norms from Colombia, Chile, México, Spain, United States, and Canada. *Archives of Clinical Neuropsychology*, 34(5), 657–681.
- Estevis, E., Basso, M. R., & Combs, D. (2012). Effects of practice on the Wechsler adult intelligence scale-IV across 3- and 6-month intervals. *The Clinical Neuropsychologist*, 26(2), 239–254.
- Fletcher, J. M., & Miciak, J. (2024). Assessment of specific learning disabilities and intellectual disabilities. *Assessment*, 31(1), 53–74.
- 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.
- Gasquoine, P. G. (2009). Race-norming of neuropsychological tests. *Neuropsychology Review*, 19(2), 250–262.
- Gasquoine, P. G. (2016). Effects of bilingualism on vocabulary, executive functions, age of dementia onset, and regional brain structure. *Neuropsychology*, 30(8), 988–997.
- Gasquoine, P. G. (2022). Performance-based alternatives to race-norms in neuropsychological assessment. *Cortex*, 148, 231–238.
- Gasquoine, P. G., Cavazos, A., Cantu, J., & Weimer, A. A. (2010). Bilingualism and Hispanic American intelligence test scores. In E. F. Caldwell (Ed.), *Bilinguals: cognition, education, and language processing* (pp. 181–199). Nova Science Publishers.
- Gasquoine, P. G., Croyle, K. L., Cavazos-Gonzalez, C., & Sandoval, O. (2007). Language of administration and neuropsychological test performance in neurologically intact Hispanic American bilingual adults. *Archives of Clinical Neuropsychology*, 22(8), 991–1001.
- Gasquoine, P. G., & Gonzalez, C. D. (2012). Using monolingual neuropsychological test norms with bilingual Hispanic Americans: Application of an individual comparison standard. *Archives of Clinical Neuropsychology*, 27(3), 268–276.
- 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.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology*, 140(2), 186–209.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67–81.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian Adults*. Psychological Assessment Resources.
- IBM Corp. (2023). *IBM SPSS statistics for windows (version 29.0)*. Computer Software.
- 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.
- Levi, J., Oppenheim, S., & Wechsler, D. (1945). Clinical use of the mental deterioration index of the Bellevue-Wechsler scale. *The Journal of Abnormal and Social Psychology*, 40(4), 405–407.
- Melendez, F. (1994). The Spanish version of the WAIS: Some ethical considerations. *The Clinical Neuropsychologist*, 8(4), 388–393.
- Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *The Quarterly Journal of Experimental Psychology Section A*, 38(4), 739–751.

- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77–101.
- Ontiveros, G., & Gasquoine, P. G. (2023). Clinical neuropsychology of bilingual Mexican American adults: Effect of language proficiency and dominance. *Archives of Clinical Neuropsychology*, acad072. Advance online publication.
- Puente, A. E., & Salazar, G. D. (1998). Assessment of minority and culturally diverse children. In A. Prifitera, & D. H. Saklofske (Ed.), *WISC—III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 227–248). Academic Press.
- Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and nan members. *Archives of Clinical Neuropsychology*, 31(3), 206–230.
- Stewart, M. (2012). *The Spanish language today*. Routledge.
- Suen, H. K., & Greenspan, S. (2009). Serious problems with the Mexican norms for the WAIS-III when assessing mental retardation in capital cases. *Applied Neuropsychology*, 16(3), 214–222.
- Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism*, 22(3), 516–536.
- U.S. Census Bureau (2023). *Median household income, 2018-2022 for Cameron and Hidalgo counties*. <https://www.census.gov/quickfacts/fact/table/cameroncountytxas,hidalgocountytxas,US/PST045223>.
- U.S. Department of Health and Human Services (2020). *Hispanic/Latino population information*. <https://minorityhealth.hhs.gov/hispaniclatino-health>.
- The Psychological Corporation (2009). *Test of premorbid functioning: Advanced clinical solutions for use with the WAIS-IV and WMS-IV*. Author.
- Wechsler, D. (1968). *Escala de inteligencia Wechsler para adultos*. Psychological Corporation.
- Wechsler, D. (2003). *WAIS-III: Escala de inteligencia de Wechsler para adultos-III*. Pearson.
- Wechsler, D. (2005). *Wechsler intelligence scale for children-fourth edition, Spanish*. Psychological Corporation.
- Wechsler, D. (2008). *Wechsler adult intelligence scale – fourth edition*. Pearson.
- Wechsler, D. (2012). *WAIS-IV. Escala de inteligencia de Wechsler para adultos-IV. Manual de aplicación y corrección*. Pearson.
- Wechsler, D. (2013). *WAIS-IV. Escala de inteligencia de Wechsler para adultos-IV. Manual de aplicación y corrección – Versión Chilena*. Pearson.
- Wechsler, D. (2014). *Escala Wechsler de inteligencia para adultos IV: Traducción al Español por editorial el manual moderno*. Pearson.
- Wechsler, D. (2016). *Escala de inteligencia de Wechsler para adultos - Cuarta edición (WAIS-IV). Manual de administración y corrección, versión estandarizada en Colombia*. Pearson.
- Weimer, A. A., & Gasquoine, P. G. (2016). Belief reasoning and emotion understanding in balanced bilingual and language-dominant Mexican American young children. *The Journal of Genetic Psychology*, 177(2), 33–43.
- Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock- Muñoz language survey-revised*. Riverside Publishing.
- World Bank (2010). *Educational attainment from the Barro-Lee educational attainment dataset*. <https://databank.worldbank.org/source/education-statistics:-Education-Attainment>.