Journal of the International Neuropsychological Society (2024), **30**, 162–171
doi:10.1017/S1355617723000310

Research Article

Disruption of the serial position effect as an early marker of Alzheimer's disease in Spanish–English bilinguals

Reina Mizrahi[1], Oona Cromheecke[2], David P. Salmon[3] and Tamar H. Gollan[4]

[1]Department of Cognitive Science, University of California, San Diego, CA, USA, [2]Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium, [3]Department of Neurosciences, University of California, San Diego, CA, USA and [4]Department of Psychiatry, University of California, San Diego, CA, USA

Abstract

Objectives: The present study examined if disruption of serial position effects in list recall could serve as an early marker of Alzheimer's disease (AD) in Spanish–English bilinguals. **Methods:** We tested 20 participants initially diagnosed as cognitively normal or with mild cognitive impairment who declined and eventually received a diagnosis of AD (decliners), and 37 who remained cognitively stable (controls) over at least 2 years. Participants were tested on the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Word List Learning Test in English or Spanish as part of an annual neuropsychological evaluation. **Results:** Compared to controls, decliners exhibited significantly reduced recall including reduced primacy scores (i.e., items recalled from the *first* three list items on Trial 1), whereas recency scores (i.e., items recalled from the *last* 3 list items on Trial 1) were equivalent in decliners and controls. Further analyses suggested that the sensitivity of the primacy effect to preclinical AD was initially stronger in participants tested in Spanish, a surprising finding given that the CERAD was developed for English speakers. However, in the subsequent year of testing, primacy scores declined to the same level regardless of language of testing. **Conclusions:** Several list learning measures may facilitate early diagnosis of AD in Spanish–English bilinguals, possibly including the relatively understudied primacy effect. Additional studies are needed to investigate the possibility that linguistic or demographic variables might modulate sensitivity of list learning tests to preclinical AD, which could lead to broader improvements in their utility for early diagnosis of AD in all populations.

Keywords: Alzheimer's disease; Bilingualism; Serial position effects; Primacy effects; List memory; Neuropsychological tests

(Received 10 November 2022; final revision 2 April 2023; accepted 26 April 2023; First Published online 21 June 2023)

Introduction

Approximately 6.5 million people in the U.S. age 65 and older have Alzheimer's disease (AD) (Alzheimer's Association, 2022). This number is expected to more than double by 2060 with a large portion of this growth projected to occur in Hispanics. There is general consensus that neural changes of AD begin prior to the observation of significant clinical symptoms (Sperling et al., 2011) and that subtle cognitive changes occur before a clinical diagnosis of dementia and probable AD can be made with any certainty. The ability to identify these early cognitive changes in a growing and diverse population is imperative as new treatment options are being developed that are likely to be most effective early in the course of AD (Cummings et al., 1994). Thus, the field has shifted to identifying neuropsychological markers of "preclinical" AD to provide earlier diagnosis to inform treatment options. This effort has identified a number of neuropsychological and neuroimaging markers of AD in the preclinical period of the disease (Twamley et al., 2006). It should be noted, however, that much of this research was performed only with majority English-speaking populations, so relatively little is known about the efficacy of cognitive markings of preclinical AD in individuals from diverse cultural backgrounds.

Cultural and linguistic factors can lead individuals to under- or over-perform on various neuropsychological tests increasing the possibility of misdiagnosis (Gasquoine & Gonzalez, 2012; Weissberger et al., 2013). Efforts have been made to identify these biases and to reduce or eliminate them by modifying cognitive tests to make them more appropriate for the target population or by employing culturally and linguistically appropriate control comparison groups (Ardila et al., 1994; Judd et al., 2009; Pedraza & Mungas, 2008; Siedlecki et al., 2010; Weissberger et al., 2013; Weissberger et al., 2017). Weissberger et al. (2013) compared neuropsychological test scores of clinically normal Hispanic and non-Hispanic elderly individuals who had cognitive decline in subsequent years (i.e., *decliners* or preclinical AD) to respective demographically appropriate cognitively healthy Hispanic or non-Hispanic control groups composed of individuals who remained stable in subsequent years. They found that cognitively stable Hispanic controls obtained lower scores on tests of language, executive function, and some measures of global cognition than non-Hispanic controls. In addition, several tests (e.g., picture naming, Trail-Making Test parts A and B) were performed worse by decliners than controls in both the Hispanic and non-Hispanic groups (marking preclinical AD), but some

Corresponding author: Reina Mizrahi; Email: rmizrahi@ucsd.edu

Cite this article: Mizrahi R., Cromheecke O., Salmon D.P., & Gollan T.H. (2024) Disruption of the serial position effect as an early marker of Alzheimer's disease in Spanish–English bilinguals. *Journal of the International Neuropsychological Society*, **30**: 162–171, https://doi.org/10.1017/S1355617723000310

language-based tests (i.e., vocabulary, semantic fluency) were sensitive to preclinical AD only in non-Hispanic groups.

Performance on language-based tests might be especially likely to vary by cultural group given the difficulty of producing translation equivalent versions of tests and the effects of bilingualism on test performance (Peña, 2007; for review, see Kroll & Gollan, 2013). On confrontation naming tests, for example, bilinguals name pictures more slowly than monolinguals, have more tip-of-the-tongue states (Gollan & Silverberg, 2001), and have higher error rates than monolinguals, even when naming pictures in their dominant language (Gollan et al., 2005; Gollan & Acenas, 2004; Ivanova & Costa, 2008; Roberts et al., 2002). Using directly translated versions of tests designed in a different language (e.g., the Boston Naming Test) may magnify differences between bilinguals and monolinguals and provide an inaccurate assessment of language proficiency and dominance (Gollan et al., 2012). Attempts to create linguistically equivalent translations may reduce, but not eliminate, this negative effect on performance depending on how equivalence is judged given idiosyncratic differences between languages that cannot be easily controlled, such as cultural interpretation, differences in word frequency and length, dialectal variation, and familiarity with the terms and/or concepts (Peña, 2007; Restrepo & Silverman, 2001).

Despite evidence that linguistic factors can reduce test sensitivity, little research has addressed the possible effects of language of testing on word list learning and memory tests. Word list learning tasks generate some of the most widely used measures for detection of preclinical AD (Weissberger et al., 2017; for review see Belleville et al., 2017), including the amount of learning across repeated presentation-immediate recall trials (i.e., hearing/reading the list and then immediately producing the words), delayed recall (i.e., producing the words after an interposed time interval filled with an unrelated cognitive task), and delayed recognition (i.e., recognizing the words after an interposed time interval filled with an unrelated cognitive task; Bondi et al., 1994). Additional features such as susceptibility to proactive interference or a reduction in semantic clustering are known to be sensitive to AD (Delis et al., 2010; Loewenstein et al., 2016; Rosselli et al., 2019) and can be derived from tests that include multiple word lists or words from multiple semantic categories (Breton et al., 2021; Delis et al., 1987; González et al., 2002; Rosselli et al., 2019).

Another aspect of performance on a word list learning test that is affected by AD is the well-established primacy effect – better immediate free recall of words from the beginning of a list compared to words positioned in the middle of the list (Atkinson & Shiffrin, 1968). The primacy effect is often attributed to greater attention to early list words and their more extensive rehearsal in long-term (episodic) memory. A recency effect also occurs with better recall for items at the end of the list than for middle items, an effect attributed to immediate production of words that are still in short-term memory at time of recall. Studies have demonstrated reduced primacy effects in individuals with mild cognitive impairment (MCI) and/or AD while recency effects remained unaffected (Bayley et al., 2000; Foldi et al., 2003; Howieson et al., 2011; Kloth et al., 2020; Moser et al., 2014). In addition, there is a negative association between primacy effects (but not recency effects) and AD pathology (Gicas et al., 2020). A reduced primacy effect may also be a marker of preclinical or prodromal AD (Kloth et al., 2020). A reduced primacy effect on the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Word List Learning Test improved prediction of future cognitive decline in individuals with MCI (Cunha et al., 2012; Moms et al., 1989). Clinically normal individuals with elevated cerebrospinal fluid levels of phosphorylated tau (a biomarker of AD) and a reduced primacy effect on the Buschke Selective Reminding Test (a 16-item word list memory and recall test) had increased risk of developing dementia related to AD, but only if they also had smaller hippocampal volume (Bruno et al., 2015).

Given the potential of the primacy effect as a cognitive marker of preclinical AD, it is important to understand the possible effects of language of testing on this feature of word list learning performance. Serial position effects (including primacy and recency effects) and their potential sensitivity to AD in Hispanics have not been examined. The present study examined serial position effects on the CERAD Word List Learning Test (Fillenbaum et al., 2007; Morris et al., 1989) in cognitively intact Hispanic decliners vs. non-decliners (i.e., controls) to determine if performance would be sensitive to future cognitive decline in this demographic. English and Spanish versions of the CERAD Word List Learning Test were used, as appropriate, which made it possible to examine language of testing effects. Previous studies have shown that English and Spanish versions of the CERAD Word List Learning Test have equivalent ability to detect memory impairment after controlling for education level (Carrión-Baralt et al., 2009; Fillenbaum et al., 2007; Ostrosky-Solís et al., 1999; but see O'Bryant et al., 2018); however, one study reported that the CERAD was less consistent in distinguishing levels of dementia severity in Spanish speaking than English-speaking participants (Fillenbaum et al., 2005). The CERAD Word List Learning Test may be well suited for examining serial position effects since it does not include semantically related words which could lead to clustering that counteracts serial position effects and is an optimal length for generating primacy effects. Previous studies in young monolingual participants have shown that primacy effects are maximized in free recall of shorter (6–8 words) compared to longer (14–15 words) word lists (Ward et al., 2010). The CERAD word list is 10 items (repeated over three learning trials) which should produce strong primacy effects.

Methods

Participants

The research protocol was approved by the UCSD Institutional Review Board in accordance with the Helsinki Declaration. Participants were recruited from the UCSD Alzheimer's Disease Research Center (ADRC). Annual ADRC evaluations included detailed clinical and medical history, brief medical examination, neurological and neuropsychological assessment, screening for depression and other psychiatric symptoms, assessment of functional activities of daily living, and laboratory tests. At the completion of the evaluation two ADRC board-certified neurologists reviewed all information (in consultation with neuropsychologists) and classified each participant as either cognitively normal or cognitively impaired, and if cognitively impaired, diagnosed them with MCI (using current NIA-AA diagnostic criteria; Albert et al., 2011), AD (using current NIA-AA diagnostic criteria; McKhann et al., 2012), or another neurodegenerative disease (e.g., frontotemporal dementia, dementia with Lewy bodies) based on published criteria. Exclusion criteria were history of stroke, severe head trauma, substance abuse, or major neurological (other than AD dementia or MCI), psychiatric or metabolic disorder. All ADRC procedures received institutional

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

Table 1. Participant characteristics

	Control ($n = 37$)		Decliner($n = 20$)		Mild AD($n = 20$)		Moderate AD ($n = 31$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	78.1	7.3	79.4	8.4	73.5	7.7	76.9	7.6
Education	12.9	3.5	11.6	3.3	13.8	3.7	10.5	4.9
Years to conversion	.	.	−1.4	0.9	.	.	.	.
Number of previous evaluations	8.8	6.2	9.0	6.0	1.0	0.0	1.0	0.0
MMSE	27.6	5.0	27.0	2.3	25.2	2.2	21.5	4.3
DRS***	134.5	6.3	127.0	5.2	125.2	3.2	107.4	9.8
Trail-Making Test, Part A[†]	44.2	14.9	52.6	16.8	63.9	27.3	82.4	41.5
Trail-Making Test, Part B***	145.7	78.7	235.5	68.0	173.3	77.5	251.8	69.5
Semantic Fluency***	42.4	11.0	32.9	7.5	31.3	9.0	24.2	8.7
Letter Fluency***	36.0	10.7	28.0	9.1	26.9	9.6	23.1	11.8
Digit Span Forward	5.5	1.1	5.7	0.9	5.6	0.8	5.2	1.0
Digit Span Backward	4.0	1.1	3.7	0.9	3.6	0.7	3.2	1.1
MINT dominant language**	63.7	5.6	59.3	5.0	59.8	5.3	58.5	4.1
MINT nondominant language	43.7	15.4	44.9	13.1	39.0	17.3	33.2	17.9
Bilingual Index Score	0.69	0.25	0.76	0.23	0.66	0.33	0.57	0.31

***$p \leq .001$, **$p \leq .01$, †=$p \leq .10$; *p*-values comparing decliners to controls (all other scores in these two columns did not differ significantly all *ps* $\geq .18$). MMSE = Mini Mental State Exam; Semantic Fluency = Animals, Fruits, Vegetables; Letter Fluency = F, A, S; Digit Span Scores = highest span correct on at least one of two trials.
Note: Trail-Making Test scores were available for 19/20 decliners and 34/37 controls.
Note: Digit Span scores were available for 15/20 decliners and 35/37 controls.
Note: MINT scores were available for 16/20 decliners and 32/37 controls; total possible correct on this test is 68; scores reflect the highest span correct out of two trials, and about most.
[a]Decliners were tested with the WAIS-R version, whereas most controls were tested with Digit Span.

ethics approval from the UCSD Human Research Protection Program and all participants provided written informed consent.

Participant recruitment

Data from 57 Hispanic participants were included in analyses: 20 who were initially diagnosed as cognitively normal or MCI but had subsequent cognitive decline and eventually received a diagnosis of AD (decliners) and 37 who remained cognitively stable (controls) over at least 2 subsequent ADRC longitudinal evaluations. Of the 20 decliners, 11 had a diagnosis of MCI in the year prior to diagnosis, 3 demonstrated memory impairment without MCI, and 6 were cognitively normal but proceeded to decline the following year. Following Weissberger et al. (2013), participants were judged to be cognitively normal based on medical and neuropsychological assessments (see Galasko et al., 1994, for more details) and decliners were identified as individuals that converted from cognitively normal to a diagnosis of probable AD. Probable AD was determined by using criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (using current NIA-AA diagnostic criteria; McKhann et al., 2012).

Decliners and controls were matched for age, education, number of prior ADRC annual evaluations they had received, and did not differ significantly in Mini Mental State Exam (MMSE) scores which were in the normal range at the evaluation of interest (i.e., the annual evaluation completed in the year prior (or as close as possible to the year prior)) to the year of diagnosis of AD. Decliners and controls were also equivalent in degree of bilingualism as measured by their relative ability to name pictures in each language on the Multilingual Naming Test (see Table 1). For this purpose, we calculated a Bilingual Index Score which divides the language with the lower naming score by the language with the higher naming score (a score of 1 represents a perfectly balanced bilingual who named the same number of pictures in each language; Gollan et al., 2010, 2012). Data from participants with mild and moderate AD from the same longitudinal study are shown in Tables 1 and 2 and in Figure 1 for visual comparison purposes only. These participants were not included in the analyses. Individuals with mild and moderate AD demonstrate a step-wise decline in overall recall and in the primacy effect across severity of AD. For example, primacy scores were lowest in participants with moderate AD, followed by mild AD, then decliners, while controls exhibited robust primacy effects.

Table 2. Performance of cognitively healthy controls, demographically matched decliners, and participants with mild or moderate AD on the CERAD

	Control(-$n = 37$)		Decline-r($n = 20$)		Mild AD(n =-20)		Moderate AD($n = 31$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trial 1 correct*	4.6	1.7	3.5	1.4	2.8	1.0	2.0	1.6
Trial 2 correct**	7.0	1.5	5.6	1.7	4.5	1.4	3.4	1.7
Trial 3 correct**	7.6	1.1	6.7	1.2	5.1	1.5	4.5	2.0
Total correct***	19.2	3.4	15.8	3.3	12.4	3.0	9.8	4.6
Delay correct***	6.2	1.7	4.3	2.1	2.5	2.4	0.8	1.5
Total Intrusions Trials 1-3*	0.7	1.1	1.5	1.9	1.7	1.9	1.7	1.9
Delay Intrusions	0.3	0.6	0.6	1.0	1.2	1.7	1.0	1.9
Trial 1 primacy*	1.6	0.9	1.1	0.8	0.8	0.8	0.5	0.7
Trial 1 middle	1.3	1.1	0.9	0.9	0.7	0.6	0.5	0.9
Trial 1 recency	1.6	0.8	1.5	0.9	1.3	0.7	1.0	0.8
Recognition Correct YES*	9.6	0.8	9.1	1.0	8.0	2.6	7.6	2.9
Recognition Correct NO[†]	9.8	0.4	10.0	0.2	9.3	1.6	8.3	1.5

***$p \leq 001$, **$p \leq .01$, *$p < .05$, †=$p \leq .10$; *p*-values comparing decliners to controls (all other scores in these two columns did not differ significantly all ps $\geq .12$).
Note: The total possible correct for primacy (list items 1–3) and recency scores (list items 8–10) is 3 while the total possible correct for the middle score is 4 (list items 4–7).

Materials and procedure

Participants were tested annually on the CERAD Word List Learning Test (Welsh et al., 1994) in English or Spanish as part of their annual neuropsychological evaluation in the UCSD ADRC longitudinal study. The test consists of immediate free recall of a

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

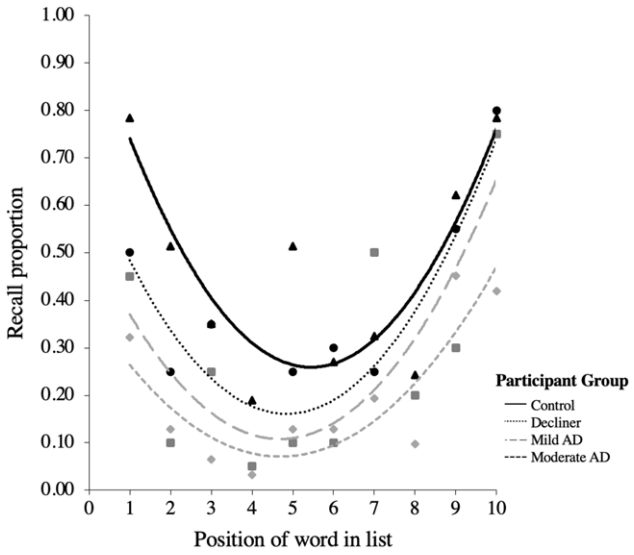

Figure 1. Trial 1 serial position curves for each participant group.

10-item word list assessed over three separate learning trials. On the first learning trial, 10 printed words are presented one at a time at a rate of one every 2 s to be read aloud by the participant. If a participant cannot read a word, it is read aloud for them by the examiner. Immediately following the presentation of the list, the participant is asked to recall as many of the words as possible – this is referred to as immediate recall. Two additional learning trials are immediately administered following the same procedure. After a delay of 5–8 min filled with unrelated testing, the participant is again asked to recall as many of the words as possible from the 10-item list (i.e., delayed recall). After delayed recall, a yes/no recognition task is administered in which the original 10 words and 10 intermixed distractor words are presented one at a time (i.e., delayed recognition).

The Spanish translation of the CERAD Word List Learning Test used in the current study was a direct translation of the English words into Spanish initiated at the ADRC in 1992, prior to the development of other Spanish–English equivalent versions of the test. According to the Subtitle-Based Word Frequency Estimates for English and Spanish (SUBTLEX-US and SUBTLEX-ESP), the English and Spanish words in our version were roughly matched at an average word frequency of 37 per million words (Brysbaert & New, 2009; Alonso et al., 2011, Cuetos et al., 2012), but Spanish words were longer (2.4 syllables), on average, than English words (1.5 syllables). In addition, Spanish translations were not always exact translations of the English word (e.g. *shore* was translated as *playa* which means *beach*), and some were cognates (i.e., translation equivalents that are similar in form) or false cognates (e.g. *engine* was translated as *motor* which would be back-translated as *motor* in English; and *pole* was translated as *polo* which in Spanish more often refers to the game than to the object). In the analyses and discussion below, we considered the possible effects of language of testing on performance.

Data analysis

Our primary aim was to identify if the primacy effect declines pre-clinically in Spanish–English bilinguals. The analyses were selected to address this question. We calculated a primacy score (see Cunha et al., 2012; Moser et al., 2014) which included the total number of items recalled from the first three words on the presented list on Trial 1 (range 0–3). We focused on Trial 1 because primacy scores flatten with increasing learning trials (as more items from the middle of the list are recalled). Using the primacy score as the dependent variable, we examined the effects of group (decliner, control) and the language used for assessment (English, Spanish) in a 2 × 2 ANCOVA with group and language of testing as fixed factors, and age and education level as continuous covariates. For comparison, we also report analyses comparing decliners and controls on total correct, delayed recall correct, and five other neuropsychological (non-memory) tests (i.e., the Dementia Rating Scale (DRS), Trail-Making Test B, Semantic Fluency, Letter Fluency, and dominant language MINT score). Data were analyzed using SPSS, version 26 (IBM; Chicago, IL). All data have been uploaded to Open Science Framework (https://osf.io/hgm3e/).

Results

Performance on measures from the CERAD Word List Learning Test is shown in Table 2. Analysis of primacy scores, controlling for age and education, revealed a main effect of group $F(1,58) = 5.97$; $\eta_p^2 = .11$; $p = .02$, and no main effect of language $F(1,58) = 1.47$; $\eta_p^2 = .03$; $p = .23$, but a significant interaction $F(1,58) = 4.54$; $\eta_p^2 = .08$; $p = .04$, showing that the difference between decliners and controls was driven by participants tested in Spanish, see Tables A1–A2, Figures 2–3 (controls tested in Spanish recalled 1.7 of the 3 primacy words on average, while decliners recalled 0.69 on average). The effects of age and education were not significant, all $ps \geq .11$.

Analysis of total correct showed a main effect of group $F(1,58) = 11.99$; $\eta_p^2 = .19$; $p = .001$, with decliners having lower total correct than controls. There is also a main effect of language, $F(1,58) = 5.02$; $\eta_p^2 = .09$; $p = .03$, with lower total correct for those tested in Spanish versus in English. The interaction between language and group was not significant, $F(1,58) = 2.64$; $\eta_p^2 = .56$; $p = .11$. In this ANCOVA, age was marginally significant, $F(1,58) = 3.04$; $\eta_p^2 = .06$; $p = .09$, and the effect of education was not significant, $F < 1$.

Analysis of delayed recall revealed a main effect of group, $F(1,58) = 10.96$; $\eta_p^2 = .18$; $p = .002$, with decliners recalling less words than controls. The main effect of language was marginally significant, $F(1,58) = 3.66$; $\eta_p^2 = .07$; $p = .06$, as was the interaction between language and group, $F(1,58) = 3.51$; $\eta_p^2 = .06$; $p = .07$. Similar to total recall, the covariate of age was marginally significant, $F(1,58) = 3.67$; $\eta_p^2 = .07$; $p = .06$, and the effect of education was not significant, $F < 1$.

All analyses comparing decliners and controls on five non-memory tests (i.e., the Dementia Rating Scale (DRS), Trail-Making Test B, Semantic Fluency, Letter Fluency, and dominant language MINT score; see Table 1) revealed main effects of group (all $ps \leq .02$), no main effects of language (all $ps \geq .38$), and no significant interactions ($ps \geq .11$[1]). Three tests produced main effects of education (DRS, $p < .001$; Trail-Making Test B, $p = .001$; dominant language MINT score, $p = .04$), and two produced significant age effects (DRS, $p = .03$; Trail-Making Test B, $p = .02$). These findings differed from those for the CERAD Word List Learning Test; specifically, the interaction between group and

[1]All $ps \geq .25$ except for the Dementia Rating Scale which exhibited a trend ($p = .11$) toward an interaction in the opposite direction as the CERAD measures, with the difference between decliners and controls being greater for participants tested in English (controls scored 135.4 on average, while decliners scored 127.3, a difference of 8.1) than for participants tested in Spanish (controls scored 132.3 on average, while decliners scored 128.4, a difference of just 3.9 points).

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

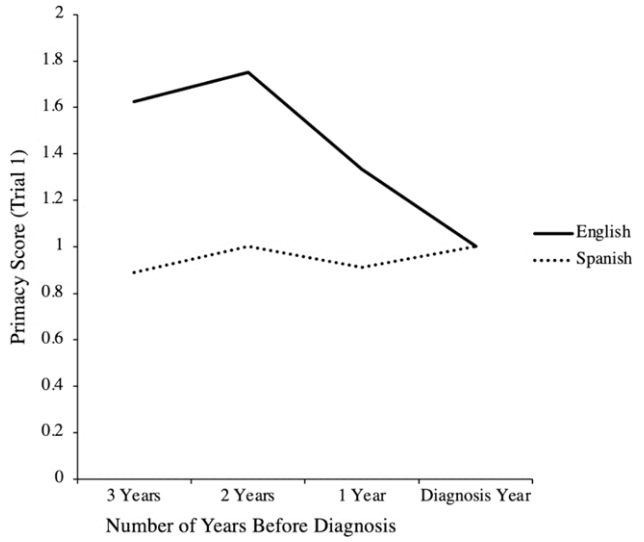

Figure 2. Decliners primacy scores for the years leading up to probable AD diagnosis.

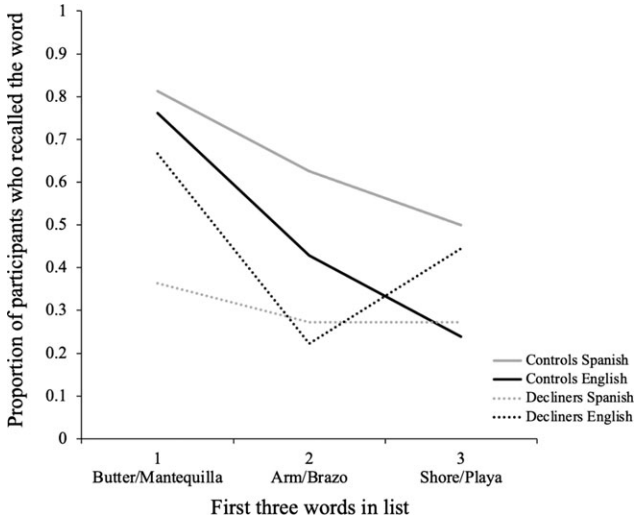

Figure 3. Proportion of participants who recalled each of the first three words on Trial 1.

language was only present for primacy scores on the CERAD Word List Learning Test and appears to reflect properties unique to this test.

Discussion

The results of the present study show that a number of measures from the CERAD Word List Learning Test, including the primacy score, were worse in decliners than in controls in the year prior to receiving a diagnosis of AD. Decliners scored worse than controls on total words recalled during learning and delayed recall consistent with previous studies that have shown subtle episodic memory decline in individuals with prodromal/preclinical AD (Twamley et al., 2006). Of particular interest, decliners exhibited a reduced primacy effect relative to controls in the presence of a preserved recency effect (see Table 2). The U-shaped serial position curves in Figure 1 illustrate a clear drop in recall for items in early but not in late list positions for decliners relative to controls. This finding replicates studies that have demonstrated the sensitivity of primacy effects to preclinical or prodromal AD (Bruno et al., 2015; Cunha et al., 2012), and extends the finding to Hispanic participants. This increases the generalizability of the result to diverse populations and boosts confidence in the utility of reduced primacy as a measure for early detection of AD.

A surprising aspect of our results was that the sensitivity of the primacy effect to preclinical AD was more prominent in participants tested in Spanish rather than English, even though the test was originally designed for administration in English. While it is common to find that translated versions of a test do not work as well as the original (as reviewed above), the opposite pattern is atypical. This result is not likely to be due to demographic or cognitive differences in the two decliner groups since groups were matched for number of years prior to diagnosis, did not differ in MMSE scores, and we controlled statistically for differences in age and education level across language groups. Longitudinal data suggested that decliners tested in English exhibited a reduced primacy effect relative to their previous years of testing (within participant), and by the time of diagnosis (the year subsequent to the year shown in Tables 1–2 and in Figure 1) their primacy scores were at the same level as those tested in Spanish (see Figure 2). This provides converging evidence for the vulnerability of primacy scores to preclinical AD but leaves open the question of why this effect occurred 1–2 years sooner in those tested in Spanish. Additional information is provided in the Appendix which shows participant demographics (Table A1), test scores (Table A2), serial position curves (Figure A1), and box plots (Figure A2) broken down by language of testing.

One possible difference that might explain the more prominent drop in the primacy effect in participants tested in Spanish rather than English might be that those tested in English completed a slightly different and longer battery of cognitive tests as part of their annual longitudinal evaluation than those tested in Spanish, potentially leading to more fatigue, and only those tested in English also completed another word list learning test (the California Verbal Learning Test) prior to the CERAD Word List Learning Test (the word list learning tests were separated by 1–2 hours filled with unrelated testing and no words overlapped on the two tests). Both groups also remembered short non-overlapping word lists for the MMSE. Another difference is that participants tested in Spanish were more likely to have a diagnosis of MCI in the year prior to diagnosis (8/11 or 73%) than were those tested in English (3/9 or 33%). This might reflect greater difficulty diagnosing AD in participants tested in Spanish (due to possible cultural, linguistic, and demographic factors) so that they were further along in their decline trajectory than those tested in English. Consistent with this possibility, decliners tested in Spanish tended to have numerically lower test scores than decliners tested in English, though none of these differences were significant, all $ps \geq .09$. However, this cannot provide a complete explanation for the language of testing differences we observed since the sharp drop in primacy scores for participants tested in Spanish was driven at least in part by unusually high primacy scores for controls tested in Spanish (see Table A2).

To further explore the effects of language of testing on CERAD Word List Test primacy effects, Figure 3 illustrates the proportion of participants in each group that recalled each of the first three words on Trial 1. This showed that controls tested in Spanish tended to recall primacy items more often than did controls tested in English (even though controls tested in Spanish had lower

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

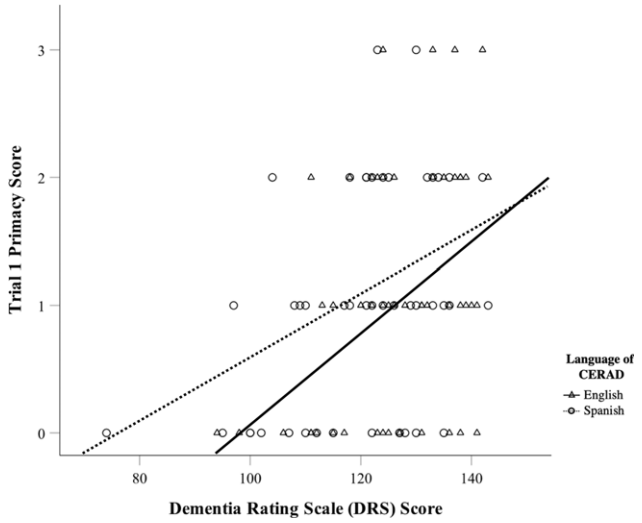

Figure 4. Primacy score on Trial 1 of the CERAD plotted against DRS score for all participants tested in each language including controls, decliners, mild AD, and moderate AD (see Table 1).

education level than controls tested in English). Thus, greater sensitivity of the primacy effect to preclinical AD in Spanish than in English seemed to be driven as much by performance of the controls as by performance of the decliners. It might seem that this could be attributed to differences in word length (Spanish words tend to be longer than English words) but note that in both languages the first word was twice as long as the second and third words (in number of syllables). As Figure 3 shows, primacy effects were evident in controls in both languages even within the first three words on the list – they recalled the first word most often and the third word least often. However, the first and best recalled word in each language was also the longest of the three words confounding serial position and word length (note that longer words are recalled better than short words when long and short words are intermixed within a list; Katkov et al., 2014). The greater sensitivity of the primacy effect in Spanish than in English was limited to the preclinical AD stage. When all participants were included in the analysis, including those with mild and moderate dementia, primacy scores were significantly correlated with scores on the DRS and these correlations were equally strong in both languages, or if anything, were stronger in English, $r = .415$, $p < .001$ than in Spanish, $r = .381$, $p = .006$ (see Figure 4).

The present results also showed that a number of neuropsychological tests besides the CERAD Word List Learning Test are sensitive to preclinical AD in elderly Hispanic individuals, and about equally so for those who prefer to be tested in English versus Spanish. Despite similar MMSE scores, decliners performed worse than controls on the Trail-Making Test B, a picture naming test (here measured with the MINT instead of the BNT as in Weissberger et al., 2013), the DRS, and semantic and letter fluency tests. In contrast to the present results, the DRS and fluency tests did not differ between Hispanic decliners and controls in Weissberger et al. (2013), a difference that might be attributable to differences in the interval between testing and conversion to probable AD: under 2 years in the present study (see Table 1) and closer to 5 years in Weissberger et al. (2013). That is, decliners in the present study were likely further along in the course of AD than those in Weissberger et al. (2013).

Additional studies will be needed to determine if and why the nature of the primacy effect varied across languages. Our results raise the possibility that certain linguistic variables may heighten sensitivity of primacy effects to preclinical AD and invite further study of the primacy effect in diverse populations. This effort seems worthwhile since the population of Hispanic elders is expected to grow seven-fold by 2050 (U.S. Census Bureau, 2004), the prevalence of AD is expected to increase exponentially (Novak & Riggs, 2004), and understanding linguistic differences will be broadly informative as to why the primacy effect is sensitive to early AD.

Funding statement. This research was supported by grants from NIA [R01 (AG076415), P50 (AG05131), P30 (AG062429)] and NSF (BCS 1923065). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NIH or NSF.

Competing interests. We have no conflicts of interest to disclose.

References

(2022). Alzheimer's disease facts and figures. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *18*(4), 700–789.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279.

Alonso, M. A., Fernandez, A., & Díez, E. (2011). Oral frequency norms for 67,979 Spanish words. *Behavior Research Methods*, *43*(2), 449–458.

Ardila, A., Rosselli, M., & Puente, A. E. (1994). *Neuropsychological evaluation of the Spanish speaker*. Springer Science & Business Media.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation (vol.* vol. *2*, p. 89–195). Academic Press.

Bayley, P. J., Salmon, D. P., Bondi, M. W., Bui, B. K., Olichney, J., Delis, D. C., Thomas, R., & Thal, L. J. (2000). Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *Journal of the International Neuropsychological Society*, *6*(3), 290–298.

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., Croteau, J., & the Consortium for the Early Identification of Alzheimer's Disease-Quebec (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(4), 328–353.

Bondi, M. W., Monsch, A. U., Galasko, D., Butters, N., Salmon, D. P., & Delis, D. C. (1994). Preclinical cognitive markers of dementia of the Alzheimer type. *Neuropsychology*, *8*(3), 374–384.

Breton, J., Stickel, A. M., Tarraf, W., Gonzalez, K. A., Keamy, A. J., Schneiderman, N., Marquine, M. J., Zlatar, Z. Z., Salmon, D. P., Lamar, M., Daviglus, M. L., Lipton, R. B., Gallo, L. C., Goodman, Z. T., & González, H. M. (2021). Normative data for the Brief Spanish-English Verbal Learning Test for representative and diverse Hispanics/Latinos:Results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), e12260.

Bruno, D., Grothe, M. J., Nierenberg, J., Zetterberg, H., Blennow, K., Teipel, S. J., & Pomara, N. (2015). A study on the specificity of the association between hippocampal volume and delayed primacy performance in cognitively intact elderly individuals. *Neuropsychologia*, *69*, 1–8.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new

and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Carrión-Baralt, J. R., Meléndez-Cabrero, J., Beeri, M. S., Sano, M., & Silverman, J. M. (2009). The neuropsychological performance of nondemented Puerto Rican nonagenarians. *Dementia and Geriatric Cognitive Disorders*, *27*(4), 353–360.

Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2012). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, *33*(2), 133–143.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*, *44*(12), 2308–2308.

Cunha, C., Guerreiro, M., de Mendonça, A., Oliveira, P. E., & Santana, I. (2012). Serial position effects in Alzheimer's disease, mild cognitive impairment, and normal aging: Predictive value for conversion to dementia. *Journal of Clinical and Experimental Neuropsychology*, *34*(8), 841–852.

Delis, D. C., Fine, E. M., Stricker, J. L., Houston, W. S., Wetter, S. R., Cobell, K., Jacobson, M. W., Salmon, D. P., Bondi, M. W. (2010). Comparison of the traditional recall-based versus a new list-based method for computing semantic clustering on the California Verbal Learning Test: Evidence from Alzheimer's disease. *The Clinical Neuropsychologist*, *24*(1), 70–79.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). California Verbal Learning Test. *Assessment*.

Fillenbaum, G. G., Kuchibhatla, M., Henderson, V. W., Clark, C. M., & Taussig, I. M. (2007). Comparison of performance on the CERAD neuropsychological battery of Hispanic patients and cognitively normal controls at two sites. *Clinical Gerontologist*, *30*(3), 1–22.

Fillenbaum, G. G., Mccurry, S. M., Kuchibhatla, M., Masaki, K. H., Borenstein, A. R., Foley, D. J., Heyman, A., Larson, E., & White, L. (2005). Performance on the CERAD neuropsychology battery of two samples of Japanese-American elders: Norms for persons with and without dementia. *Journal of the International Neuropsychological Society*, *11*(2), 192–201.

Foldi, N. S., Brickman, A. M., Schaefer, L. A., & Knutelska, M. E. (2003). Distinct serial position profiles and neuropsychological measures differentiate late life depression from normal aging and Alzheimer's disease. *Psychiatry Research*, *120*(1), 71–84.

Galasko, D., Hansen, L. A., Katzman, R., Wiederholt, W., Masliah, E., Terry, R., Hill, L. R., Lessin, P., & Thal, L. J. (1994). Clinical-Neuropathological correlations in Alzheimer's disease and related dementias. *Archives of neurology*, *51*(9), 888–895.

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.

Gicas, K. M., Honer, W. G., Wilson, R. S., Boyle, P. A., Leurgans, S. E., Schneider, J. A., & Bennett, D. A. (2020). Association of serial position scores on memory tests and hippocampal-related neuropathologic outcomes. *Neurology*, *95*(24), e3303–e3312.

Gollan, T. H., & Acenas, L.-A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 246–269. DOI: https://doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234.

Gollan, T. H., Salmon, D. P., Montoya, R. I., & da Pena, E. (2010). Accessibility of the nondominant language in picture naming: A counterintuitive effect of dementia on bilingual language production. *Neuropsychologia*, *48*(5), 1356–1366.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Bilingualism: Language and Cognition*, *4*(1), 63–83.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *15*(3), 594–615.

González, H. M., Mungas, D., & Haan, M. N. (2002). A verbal learning and memory test for English-And Spanish-Speaking older Mexican-American adults. *The Clinical Neuropsychologist*, *16*(4), 439–451.

Howieson, D. B., Mattek, N., Seeyle, A. M., Dodge, H. H., Wasserman, D., Zitzelberger, T., & Jeffrey, K. (2011). Serial position effects in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 292–299.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288.

Judd, T., Capetillo, D., Carrion-Baralt, J., Marmol, L. M., Miguel-Montes, L. S., Navarrete, M. G., Puente, A. E., Romero, H. R., Valdes, J., and the 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.

Katkov, M., Romani, S., & Tsodyks, M. (2014). Word length effect in free recall of randomly assembled word lists. *Frontiers in Computational Neuroscience*, *8*, 129.

Kloth, N., Lemke, J., Wiendl, H., Meuth, S. G., Duning, T., & Johnen, A. (2020). Serial position effects rapidly distinguish Alzheimer's from frontotemporal dementia. *Journal of Neurology*, *267*(4), 975–983. DOI: https://doi.org/10.1007/s00415-019-09662-w

Kroll, J. F., & Gollan, T. H. (2013). Speech planning in two languages. In *The Oxford Handbook of Language Production*

Loewenstein, D. A., Curiel, R. E., Greig, M. T., Bauer, R. M., Rosado, M., Bowers, D., Wicklund, M., Crocco, E., Pontecorvo, M., Joshi, A. D., Rodriguez, R., Barker, W. W., Hidalgo, J., Duara, R. (2016). A novel cognitive stress test for the detection of preclinical Alzheimer disease: Discriminative properties and relation to amyloid load. *The American Journal of Geriatric Psychiatry*, *24*(10), 804–813.

McKhann, G. M., Albert, M. S., & Sperling, R. A. (2012). Changing diagnostic concepts of Alzheimer's disease. In *Alzheimer's disease-modernizing concept, biological diagnosis and therapy*. vol. *28*, p. 115–121). Karger Publishers.

Moms, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*(9), 1159–1159.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*, 1159–1165.

Moser, B., Deisenhammer, E. A., Marksteiner, J., Papousek, I., Fink, A., & Weiss, E. M. (2014). Serial position effects in patients with mild cognitive impairment and early and moderate Alzheimer's disease compared with healthy comparison subjects. *Dementia and Geriatric Cognitive Disorders*, *37*(1-2), 19–26.

Novak, K., & Riggs, J. A. (2004). *Hispanics/Latinos and Alzheimer's disease*. Alzheimer's Association.

O'Bryant, S. E., Edwards, M., Johnson, L., Hall, J., Gamboa, A., & O'jile, J. (2018). Texas Mexican American adult normative studies: Normative data for commonly used clinical neuropsychological measures for English-and Spanish-speakers. *Developmental neuropsychology*, *43*(1), 1–26.

Ostrosky-Solís, F., Ardila, A., & Rosselli, M. (1999). NEUROPSI: A brief neuropsychological test battery in Spanish with norms by age and educational level. *Journal of the International Neuropsychological Society*, *5*(5), 413–433.

Pedraza, O., & Mungas, D. (2008). Measurement in cross-cultural neuropsychology. *Neuropsychology Review*, *18*(3), 184–193.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255–1264.

Restrepo, M. A., & Silverman, S. W. (2001). Validity of the Spanish Preschool Language Scale – 3 for use with bilingual children. *American Journal of Speech-Language Pathology*, *10*(4), 382–393.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*(4-6), 635–645.

Rosselli, M., Loewenstein, D. A., Curiel, R. E., Penate, A., Torres, V. L., Lang, M., Greig, M. T., Barker, W. W., Duara, R. (2019). Effects of bilingualism on verbal and nonverbal memory measures in mild cognitive impairment. *Journal of the International Neuropsychological Society*, *25*(1), 15–28.

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

Siedlecki, K. L., Manly, J. J., Brickman, A. M., Schupf, N., Tang, M. X., & Stern, Y. (2010). Do neuropsychological tests have the same meaning in Spanish speakers as they do in English speakers? *Neuropsychology*, *24*(3), 402–411.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., C. R. Jack Jr., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 280–292.

Twamley, E. W., Ropacki, S. A. L., & Bondi, M. W. (2006). Neuropsychological and neuroimaging changes in preclinical Alzheimer's disease. *Journal of the International Neuropsychological Society*, *12*(5), 707–735.

US Census Bureau (2004). *American factfinder (Vol. 3)*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1207–41.

Weissberger, G. H., Salmon, D. P., Bondi, M. W., & Gollan, T. H. (2013). Which neuropsychological tests predict progression to Alzheimer's disease in Hispanics? *Neuropsychology*, *27*(3), 343–355.

Weissberger, G. H., Strong, J. V., Stefanidis, K. B., Summers, M. J., Bondi, M. W., & Stricker, N. H. (2017). Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(4), 354–388.

Welsh, K. A., Butters, N., Mohs, R. C., Beekly, D., Edland, S., Fillenbaum, G., & Heyman, A. (1994). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology*, *44*(4), 609–609.

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

Appendix Data split by language of CERAD administration

Table A1. Characteristics of cognitively healthy controls and demographically matched decliners split by language

	English				Spanish			
	Control (n = 21)		Decliner (n = 9)		Control (n = 16)		Decliner (n = 11)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	79.9	7.1	83.2	6.7	75.8	7.1	76.3	8.7
Education	14.8	2.6	13.8	3.2	10.4	3.0	9.8	2.3
Years to conversion	.	.	−1.2	.4	.	.	−1.5	1.2
Number of previous evaluations	8.0	5.1	10.6	6.0	9.8	7.4	7.6	6.0
DRS	137.3	3.9	127.6	5.1	130.9	7.1	126.5	5.6
MMSE	28.5	1.9	27.3	2.5	26.4	7.5	26.7	2.2
Trail-Making Test, Part A	42.7	14.6	55.4	20.6	46.4	15.5	50.0	13.1
Trail-Making Test, Part B	126.5	65.8	241.8	72.6	171.3	89.0	229.9	67.1
Semantic Fluency	42.0	10.0	33.4	10.7	40.2	12.3	32.5	3.7
Letter Fluency	35.0	12.0	27.4	12.1	37.3	8.7	28.5	6.3
Digit Span Forward	6.0	1.1	6.1	1.1	4.8	.7	5.4	.5
Digit Span Backward	4.1	1.1	3.7	1.3	3.9	1.0	3.6	.5
MINT dominant language	64.8	7.1	58.9	6.5	62.3	2.2	59.8	3.2
MINT Nondominant language	43.5	17.5	49.3	5.4	44.0	12.8	40.6	17.1

Table A2. Performance of cognitively healthy controls and demographically matched decliners split by language

	English				Spanish			
	Control (n = 21)		Decliner (n = 9)		Control (n = 16)		Decliner (n = 11)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trial 1 correct	4.5	1.6	4.2	1.4	4.8**	1.8	2.9	1.1
Trial 2 correct	7.0	1.5	6.2	1.9	7.0**	1.6	5.1	1.4
Trial 3 correct	7.8†	1.3	6.9	1.2	7.4*	1.0	6.5	1.3
Total correct	19.2	3.4	17.3	3.8	19.2***	3.4	14.5	2.2
Delay correct	6.1	1.9	5.2	2.5	6.3***	1.5	3.6	1.5
Total intrusions trials 1-3	0.6*	1.0	2.0	2.2	0.8	1.2	1.1	1.6
Delay intrusions	0.2†	0.7	0.9	1.2	0.4	0.5	0.4	0.8
Trial 1 primacy	1.4	1.0	1.3	0.9	1.9***	0.7	0.9	0.7
Trial 1 middle	1.4	1.0	1.1	0.9	1.1	1.2	0.6	0.8
Trial 1 recency	1.6	0.8	1.8	1.2	1.7	0.7	1.4	0.7
Recognition correct YES	9.5	0.9	9.4	0.7	9.7**	0.6	8.7	1.1
Recognition correct NO	9.8	0.4	9.9	0.3	9.8†	0.4	10.0	0
Bilingual Index Score	0.7	0.3	0.8	0.2	0.7	0.2	0.7	0.3

***$p \leq 001$, **$p \leq .01$, *$p < .05$, †=$p \leq .10$; p-values comparing decliners to controls within language.

Note: All group comparisons should be interpreted with caution given group sizes.

Note: The total possible correct for primacy (list items 1–3) and recency scores (list items 8–10) is 3 while the total possible correct for the middle score is 4 (list items 4–7).

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press

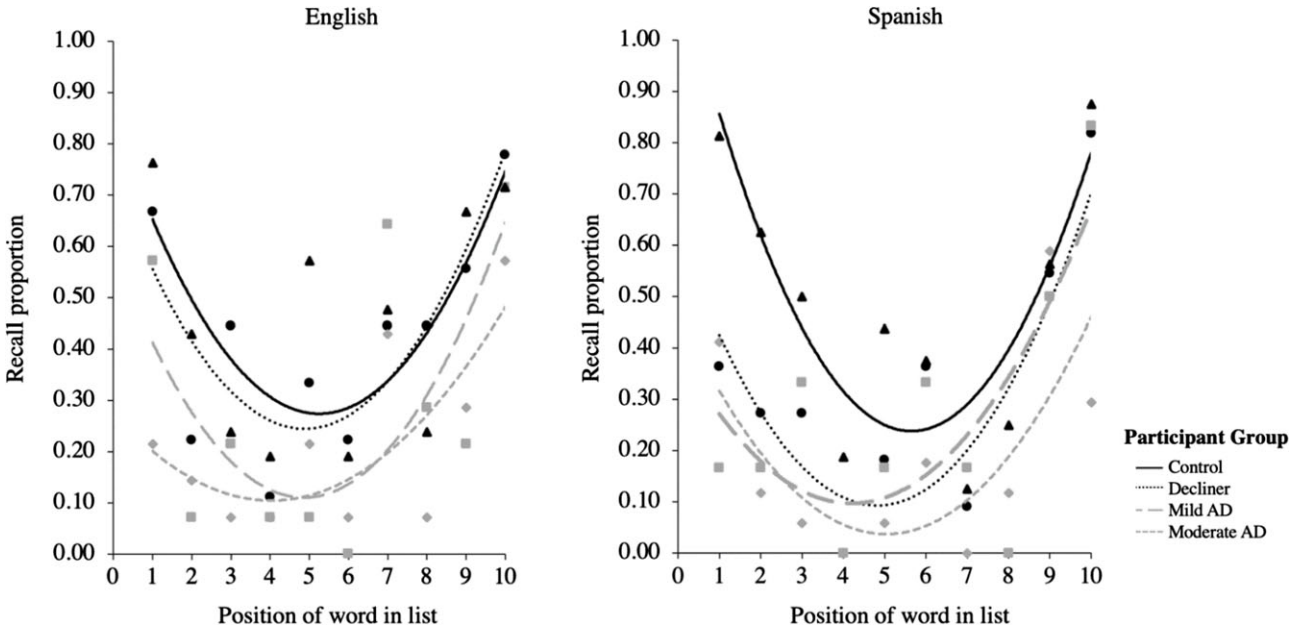

Figure A1. Trial 1 serial position curves for each participant group split by language.

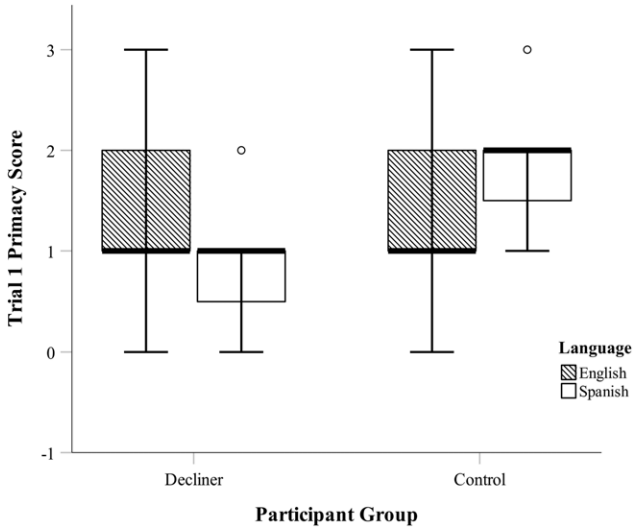

Figure A2. Trial 1 box plots split by group and language.

https://doi.org/10.1017/S1355617723000310 Published online by Cambridge University Press