Neuropsychological assessments of cognitive aging in monolingual and bilingual older adults

Article *in* Journal of Neurolinguistics · August 2016

DOI: 10.1016/j.jneuroling.2016.08.001

John A. E. Anderson
York University

Somayya Saleemi
Baycrest

Neuropsychological Assessments of Cognitive Aging in Monolingual and Bilingual Older Adults

John A. E. Anderson, Somayya Saleemi, and Ellen Bialystok

York University

Address for Correspondence:

Ellen Bialystok

Department of Psychology

York University

4700 Keele St.

Toronto, ON, Canada M3J 1P3

Email: ellenb@yorku.ca

Word count = 6,281

Acknowledgements:

This research was supported by grant R21 AG048431 from the National Institutes of Health to Ellen Bialystok.

Abstract

Standardized neuropsychological tests are routinely used as diagnostic criteria in aging populations and are an important piece of evidence for the identification of clinical pathology and neurodegenerative conditions such as Alzheimer's disease. Tests include such measures as the Mini Mental Status Exam, Delis-Kaplan Executive Function System, Montreal Cognitive Assessment, and others. These tests cover a range of functions including working memory, verbal fluency, prospective memory, and task switching. Interpretation of test results is based on comparison of the participant's score to standard scores that have been normed on a population database. However, a growing body of research has shown that the skills underlying these tests may be significantly different in monolingual and bilingual older adults, especially for those experiencing cognitive impairment, yet the standardized test scores do not account for such differences. Therefore, results of neuropsychological tests may be different for bilingual populations than for monolinguals, and those differences may be misinterpreted. The issue is important because the consequences of these interpretative errors may be over- or under-diagnosis of cognitive impairment. The present study examined the neuropsychological test scores of monolingual and bilingual older adults who were experiencing healthy aging or cognitive impairment to establish patterns in these scores that can more accurately guide the interpretation for bilingual older adults by considering group differences in the underlying abilities.

Neuropsychological Assessments of Cognitive Aging in Monolingual and Bilingual Older Adults

It has become axiomatic to begin discussions of cognitive aging with the observation that there is a significant demographic shift towards an older population. This increase in the aging population impacts all aspects of life and carries enormous consequences for personal well-being and independence, the economic conditions of individuals and societies, and health care resources as they relate to both personal and national priorities. Understanding these aging processes, therefore, is a vital precondition to planning.

Cognitive aging has a special status in the array of factors included in discussions of aging because of its direct relation to the ability of an individual to live independently. Accordingly, there is great interest in discovering approaches to maintaining cognitive function in older age with the hope of delaying or preventing dementia and the neurodegenerative diseases that accompany it. To this end, considerable effort has been placed in developing new pharmacological treatments for dementia in general and Alzheimer's disease (AD) in particular; Zhu et al. (2013), for example, conducted a large-scale multi-site study on the effect of two classes of drugs commonly used in the treatment of AD, cholinesterase inhibitors and memantine, and reported moderate success for both in prolonging life. The authors note, however, that the results are not simple in that these effects depended on a wide range of patient characteristics that influenced the outcomes. Moreover, other studies investigating the same two drug therapies have found no benefit in prolonging life. Lopez et al. (2009), for example, studied these same two drug interventions and reported no evidence for differences in life expectancy but did note that there was a delay in time until nursing home admission. Thus, the results of such pharmacological studies are highly variable. Reviews of this literature also show that even when beneficial effects are found, the efficacy of these therapies for prolonging life, including studies of cholinesterase inhibitors and memantine, are modest and typically yield small effect sizes (Massoud & Gauthier, 2010; Rockwood, 2004).

In the absence of an effective and reliable pharmacological treatment for Alzheimer's disease and other dementias, attention has turned to the lifestyle activities that have been shown to maintain cognitive function in older age. These activities, collectively known as cognitive reserve (Stern, 2002), include education, occupational status, socio-economic class, aerobic exercise and involvement in physical, intellectual and social activities (Stern et al., 1994; Bennett et al., 2003; Bennett, Schneider, Tang, Arnold, & Wilson, 2006). However, a major gap in this research is an understanding of the mechanism by which protection due to these activities takes place (Stern, 2012). The two main categories of explanation are based on the notions of brain reserve, in which more resilient brains resist neuropathology (e.g., Landau et al., 2012; Valenzuela, Sachdev, Wen, Chen, & Brodaty, 2008), and cognitive reserve, in which intact brain functions compensate for the activities of impaired ones (e.g., Bennett et al., 2006). This issue remains unresolved, but it is most likely the case that both of these mechanisms or some interaction between them are necessary for effective cognitive reserve (see Stern, 2012, for discussion). The implication is that cognitive reserve activities have consequences for brain structure, brain function, and cognitive performance.

With substantial evidence for the importance of cognitive reserve in the preservation of cognitive function in healthy aging and dementia but little understanding of the mechanism by which this protection takes place, it is important to have a full picture of the types of activities that lead to cognitive reserve. One such activity is bilingualism. Lifelong bilinguals show better cognitive function in older age than comparable monolinguals (e.g., Bialystok, Craik, Klein, &

Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013; for review, Baum & Titone, 2014). Many studies with younger adults have not shown these effects (e.g., Paap & Greenberg, 2013), a discrepancy that has been discussed elsewhere (Bak, 2016; Bialystok, in press; Kroll & Bialystok, 2013). However, with a few exceptions (e.g, Kirk, Fiala, Scott-Brown, & Kempe, 2014; Kousaie & Phillips, 2012), the results with older adults are more consistent in their outcome.

More importantly than performance on these tasks, bilinguals demonstrate symptoms of dementia at a significantly older age than monolinguals (e.g., Alladi et al., 2013; Craik, Bialystok, & Freedman, 2010; for review, Bak & Alladi, 2014). This delay of symptoms is consistent with the notion of cognitive reserve: "Individuals with high cognitive reserve, by definition, will present with disease-related clinical symptoms later than individuals with low cognitive reserve" (Stern, 2012, p. 1009). Moreover, Schweizer, Ware, Fischer, Craik, and Bialystok (2012) demonstrated that for matched groups of monolingual and bilingual patients who had been diagnosed with Alzheimer's disease but were equivalent on all clinical and neuropsychological measures, the bilingual group had significantly more disease pathology than the monolinguals. This, too, is a criterion for the identification of cognitive reserve: "…at any given level of clinical severity in Alzheimer's disease, the degree of pathology will be greater in individuals with higher cognitive reserve than in those with lower cognitive reserve" Stern (2012, p. 1008). By these criteria, therefore, bilingualism satisfies the requirements as a source of cognitive reserve.

The importance of establishing that bilingualism is a source of cognitive reserve is that it shifts the expectations for cognitive performance for groups that have this protection compared to a similar group without reserve. From the perspective of cognitive aging, this is a good thing because it means that higher levels of cognitive function are expected to be maintained with aging in the high reserve group. Thus, bilingual older adults typically outperform monolinguals on a range of cognitive tasks that are generally used to assess executive functioning, such as the Stroop task (e.g., Bialystok,Craik, & Luk, 2008a), and Simon task (e.g., Bialystok et al., 2004; Bialystok et al., 2005) among others (Baum & Titone, 2014). These tasks typically include conflict or require attending to target information in the context of misleading distraction. However, bilinguals also perform more poorly than monolinguals on verbal tasks, a trend found across the lifespan (Bialystok, 2009), specifically for tasks requiring language production, naming, or fluency (Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan, Montoya, & Werner, 2002; Kohnert, Hernandez, & Bates, 1998; Roberts, Garcia, Desrochers, & Hernandez, 2002). For this reason, tasks that assess executive function but are based on verbal stimuli show a disadvantage for bilinguals; the same tasks using nonverbal stimuli are typically performed better by bilinguals (Wodniecka, Craik, Luo, & Bialystok, 2010).

From the perspective of clinical diagnosis, however, the situation is more complex. Bilingual patients who perform equivalently to monolinguals on neuropsychological measures that are used for diagnosis turn out to have more advanced disease (Schweizer et al., 2012), meaning that the disease had gone undetected for some time. Therefore, the standard measures used in neuropsychological testing are insufficiently sensitive to small changes in cognitive level from a group of older adults with high cognitive reserve. The situation is complicated by the reliance of many of these tests on verbal ability, an area in which bilinguals are weaker than monolinguals. The combination of possibly better executive function and poorer verbal function may mask indications of cognitive impairment for bilingual older adults.

The problem in using standardized tests that have been normed on monolingual populations for the assessment of bilinguals has been known for a long time in the developmental literature. The clinical issue for children is the difficulty of diagnosing language impairment as distinct from the relatively delayed or different language acquisition patterns shown by children learning two languages. Thus, standard assessment approaches based on normalized data can both under- and over-diagnose clinical problems for bilingual children (Bedore & Peña, 2008; Paradis, Crago, Genesee, & Rice, 2003). For this reason, several initiatives have been undertaken to develop assessment tools that are appropriate for bilingual children. The problem, however, is that such instruments must be specialized for the precise two languages that the child speaks (e.g., Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore (2014) for Spanish-English bilinguals) and may have little or no generalization to bilingual children more broadly. Non-verbal tests such as the Cattell Culture Fair (Cattell & Cattell, 1973) test may be more appropriate, yet even this does not avoid the fact that the entire testing context for bilinguals is usually in the non-dominant language.

The same issues apply to the use of standardized neuropsychological measures to assess cognitive status in older adults. Rivera-Mindt et al. (2008) discuss this problem and make specific proposals for the accurate assessment of older bilingual adults that include the combination of subjective and objective measures in the assessment. Their discussion focuses primarily on Spanish-English bilinguals, but their approach could be applied more broadly to the extent that specific language tests were available in the other languages. Nonetheless, the majority of neuropsychological testing and clinical evaluation of older adults proceed with little regard for the language history of the individual or the specific changes in cognitive and linguistic performance that follow from bilingualism.

The purpose of the present study is to evaluate the performance of monolingual and bilingual older adults on neuropsychological assessments routinely used in clinical practice to identify the presence of cognitive impairment or dementia. The measures included in this study are the Mini-Mental State Exam (MMSE) and three tests from the Delis-Kaplan Executive Function System Tests (D-KEFS). In both cases, standard cut-off criteria along with neuropsychological assessment by a clinical neuropsychologist are used to determine the line between healthy cognitive aging and cognitive impairment in the clinical populations. These tests were administered to monolingual and bilingual older adults who were experiencing healthy aging (in that they had not received any diagnosis of clinical impairment) or had been diagnosed with mild cognitive impairment (MCI) or AD. Although scores from similar tests may have contributed to the clinical diagnosis made for these patients prior to the present study, those scores were not included in the present analysis. The purpose was to evaluate performance on these measures for individuals who had been designated as belonging to one of the three cognitive status groups.

The essential point about cognitive reserve is that it disrupts the usual relation between cognitive level and the demographic (e.g., age), experiential (e.g., education), or biological (e.g., brain status) variables with which it is associated. Thus, individuals with high cognitive reserve would be expected to obtain higher scores on cognitive tests than would be predicted by their age or the condition of their brains. The problem is how to describe these individuals and compare them to those with low reserve. Using cognitive measures, the high reserve individuals may demonstrate better performance than low reserve individuals, but using neurological indices, the high reserve individuals may show more compromised brain structure than low reserve individuals. Should clinical classification be based on cognitive performance, brain structure, or both? The question of diagnosis becomes more important when an entire group experiences higher cognitive reserve than their contemporaries because of a shared experience. Compounding the problem, the group, in this case, bilinguals, generally performs better than monolinguals on some nonverbal cognitive tasks but poorer than monolinguals on most verbal tasks. Since both types of tasks are normally included in assessments of cognitive status, the performance of bilinguals is difficult to interpret.

The interpretation of standard scores from neuropsychological assessment when dealing with groups who differ in cognitive reserve requires understanding what needs to be held constant or matched and what is then allowed to vary. Consider a case in which monolingual and bilingual older adults obtained similar scores on a cognitive measure. All else being equal, one would expect that individuals in these groups are of comparable age and have comparable brain status. However, either of these variables may be systematically different as a function of cognitive reserve. First, if we match for cognitive score (e.g., performance on the Stroop task) and then examine their brains, the bilinguals may have more deterioration in brain structure than the monolingual but have compensated through other means to maintain higher performance than would be expected, resulting in scores similar to the low reserve monolinguals. Similarly, if we match for cognitive assessment level (e.g. diagnosis of dementia or MCI), it may be that the bilingual group is older than the monolinguals. Two cases illustrate this point. First, in the study by Schweizer et al. (2012), monolingual and bilingual AD patients were matched on age and cognitive neuropsychology scores; the results showed that the bilinguals had more brain deterioration than monolinguals. Second, in retrospective studies of the onset of AD, monolinguals and bilinguals are essentially equivalent in neuropsychological testing in that they have crossed the clinical threshold for dementia and have similar MMSE scores, (e.g., Alladi et al., 2013), but the bilinguals are older indicating that cognitive level has been protected.

In contrast, if we match for brain integrity, then it may be the case that the bilingual group would obtain better cognitive outcomes than those in the monolingual group. Although this manipulation is less common in clinical studies, one could argue that the research with healthy aging which is based on the presumption of equivalent levels of brain integrity illustrate this point by showing better performance by bilingual than monolingual older adults (e.g., Bialystok et al., 2004). Another example comes from a recent study investigating cognitive recovery following stroke in monolingual and bilingual patients (Alladi et al., 2016). Over 600 patients who had suffered ischemic stroke and therefore were somewhat equivalent in terms of brain integrity were evaluated for the likelihood of full cognitive recovery over time following therapeutic intervention. The results showed that about 40% of the bilingual patients but only about 20% of the monolingual patients regained pre-morbid cognitive levels, all else being equal. In all these examples, the interpretation of neuropsychological test scores and their use in the assessment of cognitive level is complicated by the presence of cognitive reserve.

In the present study, we examined the performance of monolingual and bilingual older adults at three cognitive status levels – healthy aging, MCI, and AD – on standardized neuropsychological tests. The healthy aging group was a new sample collected for the present study, but the two clinical groups were obtained from a previous study in which these same assessments were used (Bialystok, Craik, Binns, Ossher, & Freedman, 2014). There were two questions examined in these data. The first was to compare the performance of monolingual and bilingual participants in each of the three cognitive status groups to determine whether these standardized tests produce valid assessments of dementia in bilingual individuals. The second was to track changes in performance on these measures across the three cognitive status levels for each of the language groups to determine how assessment changes with increasing impairment.

Method

Participants

Data were examined from 184 participants. The sample included 35 older adults experiencing healthy aging (HA), 74 individuals diagnosed with MCI, and 75 patients diagnosed with probable AD. The mean age and distribution by language group for each cognitive status level are reported in Table 1. Further details about participants in the two patient groups are reported in Bialystok et al. (2014). Participants in the patient groups had received a consensus diagnosis of MCI or probable AD from a team comprised of at least two physicians (neurologist, geriatrician, or psychiatrist) and a neuropsychologist. The instruments and results reported in the present study were not part of the diagnostic process. The HA group was recruited from the community and reported that they were not experiencing cognitive problems and had never been diagnosed with a memory or cognitive impairment.

The procedures for classifying participants by language group in the MCI and AD samples are described in Bialystok et al. (2014). For HA, participants were classified as monolingual or bilingual based on their scores on the Language and Social Background Questionnaire (LSBQ; Luk & Bialystok, 2013). Second-language proficiency level was calculated as the average score (out of 10) of questions from the LSBQ that assessed proficiency in speaking and understanding a second language and was significantly higher in bilinguals (M = 7.98, SD = 1.87) than monolinguals (M = 1.83, SD = 2.34), difference of the means = -6.3, 95% credible interval [-8.6, -4], posterior probability that the difference of the means is < 0 = 0.99. Similarly, second-language usage was the average of LSBQ questions assessing second language speaking and listening frequency; this too was higher in the bilinguals (M = 2.71, SD = 2.14) than the monolinguals (M = 0.26, SD = 0.34), difference of the means = -2.4, 95% credible interval [-3.9, -0.92], posterior probability that the difference of the means is < 0 = 0.99. Individuals were matched with monolinguals at a group level on demographic variables such as age and level of education (see Table 1). Ten bilinguals and 4 monolinguals in the HA group were immigrants, but they did not differ in age at which they arrived in Canada, (average age 27 and 33 years, respectively) with a mean difference = 5.5, 95% credible interval [-26, 39]. All testing was conducted in English. There was a very large number of non-English languages represented in the bilingual groups and it would have been impossible to conduct testing in those languages.

Tasks and Instruments

Mini-Mental State Examination *(MMSE; Folstein, Folstein & McHugh, 1975):* The MMSE is frequently employed as a basic screening measure for cognitive impairment in older adults and is commonly used in clinical settings to assess changes in patients who suffer dementia or memory loss. Although the construct validity and reliability of the test is high, the MMSE has some disadvantages (Tombaugh & McIntyre, 1992). Previous literature has shown that there is a considerable bias toward verbal items (Lancu & Olmer, 2006; Tombaugh & McIntyre, 1992; Starr, 2010). Age, education, and sociocultural background can all affect an individual's performance and test score (Strauss, Sherman, & Spreen, 2006). The language of testing in bilingual individuals can impact MMSE scores. A study by Chaoimh, De Bhaldraithe, O'Malley, Bhuí and O'Keeffe (2015) showed that bilingual English-Irish speakers scored higher when given the MMSE in Irish, their native language. The insensitivity of the MMSE is another limitation; it is useful for coarse categorization of more severe cases of cognitive impairment but ineffective for detecting lower levels of impairment. Nonetheless, the test is widely used in both the neuroscience research and medical communities as a standard assessment tool. The test is short and easily administered, both factors that have no doubt contributed to its longevity.

Delis-Kaplan Executive Function System Tests (D-KEFS; Delis, Kaplan, & Kramer, 2001). Participants completed three tests from this instrument. The D-KEFS battery is designed for use across a wide range of ability, and has excellent sensitivity to subtle impairment.

The first test was the trail making test which assesses flexible thinking in the visuomotor domain; subtests included the number sequencing (Trails A) and number–letter switching (Trails B) conditions.

The second task was the verbal fluency test, which included the subtests for letter fluency, category fluency, and category switching. Letter fluency tested participant's ability to generate words based on an initial phoneme with strict restrictions on selection and category fluency tested access to instances of conceptual categories. Category switching measured the ability to switch between two categories. There were two measurements associated with category switching. The first, total correct, was the number of correct items generated across both categories. The second, total switching accuracy, was the number of times the participant successfully switched between categories.

The third test was the color-word interference test, also known as the Stroop task. This task measures the ability to override an automatic or highly learned response. The subtests were color naming, word reading, inhibition (the Stroop effect), and inhibition-switching. A Stroop effect score was calculated from the scores obtained for naming the ink color in the standard color naming condition and naming the ink color in the inhibition condition where the color word interfered. In the inhibition-switching condition, some words were enclosed in rectangles.

Participants were instructed to *name* the ink color if there was no rectangle (same as in the Inhibition subtest) but *read* the word if it was inside a rectangle. This requires participants to switch between two rules within the same task and can reveal deficits in cognitive flexibility even if the participant has relatively intact verbal inhibition.

Procedure

Participants in the MCI and AD groups were visited in their homes by a trained research assistant. AD participants were accompanied by a family member or caregiver. Participants in the HA group came into the lab for testing. They were informed about the nature of the study, the time commitment, and the types of tasks that would be administered. A trained research assistant at York University obtained informed consent and administered the test battery. Testing occurred in a single session lasting approximately one and a half hours after which participants were compensated for their time. Test order was LSBQ, MMSE, and D-KEFS tests (trail making test, verbal fluency test, color-word interference test). In all cases, standard scores are reported. These scores are based on a population mean of 10 with a standard deviation of 3, with higher scores indicating better performance for all measures.

Results

Effect of Bilingualism within Cognitive Status

For the HA group, monolinguals and bilinguals were equivalent in age (mean difference between groups =0.09 years) and education (mean difference between groups = -0.04). In both cases, 95% credible intervals crossed zero by a similar margin on both sides. Descriptive statistics and scores for the background variables, MMSE and D-KEFS battery are reported in Table 1. MMSE scores were equivalent for the two groups and no further analyses were conducted. Standard scores from each of the D-KEFS measures were averaged to produce composite scores, one for each of the three tests. The Stroop composite was the average of color naming, word reading, inhibition, and inhibition/switching. The verbal fluency composite was the average of letter fluency, category fluency, and both category-switching measurements. Finally, the composite trail making score was the average of number sequencing and switching.

Scores for monolingual and bilingual HA on the D-KEFS battery were compared using Bayesian t-tests, as these analyses are robust to small sample sizes and outliers (Kruschke, 2013). For Bayesian comparisons we report posterior estimates of differences of the mean along with 95% credible intervals and a posterior probability. Classic 95% confidence intervals express the notion that 95 times out of 100 the true mean will be captured within the bounds of the interval. Bayesian credible intervals represent the actual probability of observing the values reported with values closer to the center of the distribution being more probable. A credible interval on a mean difference that excludes zero is evidence for a group difference (classic "significance"). Conversely, a credible interval that includes zero is evidence for a null difference if the interval is deemed sufficiently narrow. Both groups performed equally well on all measures of the trail making test, but measures from the Stroop and verbal fluency tests revealed group differences. For the Stroop task, monolinguals had a > 99% probability of outperforming bilinguals on four measures, including color naming, inhibition, inhibition switching, and inhibition errors (i.e., credible intervals excluded zero in each case). For verbal fluency measures, monolinguals had a > 98% probability of outperforming bilinguals on four measures, including category fluency, category switching (both variants), and percent repetition errors (See Table 2). Both the Stroop and verbal fluency tasks rely on verbal ability and in both cases, bilinguals produced lower scores than monolinguals. Nonetheless, the mean scores for the bilinguals were in the normal population range of 10, but the monolingual performance was higher.

To test the interpretation that Stroop scores rely on verbal ability, we investigated the relation between them (see Figure 1). An independent measures *t*-test confirmed that monolinguals had higher Stroop scores than bilinguals, $t(29.24) = 3.05$, $p = 0.005$. This analysis was followed by an ANCOVA predicting Stroop scores with group membership and verbal fluency. The model confirmed that increasing verbal fluency predicted increasing Stroop performance, $F(1,31) = 4.59$, $p = 0.04$, $\eta_p^2 = 0.13$. Neither the main effect of language group, $F(1, 31) = 0.032$, p = 0.86, nor the interaction term, $F(1, 31) = 0.012$, $p = 0.91$) were significant. These results support the interpretation that the lower performance by bilinguals on the Stroop task is related to lower verbal fluency scores of bilinguals.

Scores for the MCI and AD groups are also reported in Table 1. Previous versions of this dataset reported by Bialystok et al. (2014) used data trimming procedures and included information from multiple sessions. In contrast, the results presented in the current paper are raw, untrimmed data from only the first session to be comparable to the HA group for whom only one session was conducted. There were no significant differences in performance between monolinguals and bilinguals on any measure in the MCI group (Table 2). For AD, there were two differences between monolinguals and bilinguals for which the 95% credible interval did not include zero (Table 2). First, monolinguals (*M*= 7.75, *SD* = 4.33) were ~98% more likely to have higher letter fluency scores than bilinguals (*M*= 5.74, *SD* = 3.77), posterior mean difference = 2.1, 95% credible interval = [0.059, 4.2], consistent with better verbal ability in monolinguals. Second, bilinguals (*M*= 5.81, *SD* = 4.10) were ~99% more likely to have higher scores on inhibition/switching errors than monolinguals (*M*= 2.81, *SD* = 3.49) on the Stroop task, posterior mean difference = -4.70, 95% credible interval = [-7.0, -2.5].

Changes in Performance across Cognitive Status

The previous analyses compared performance of monolinguals and bilinguals within each cognitive status level. The next analysis examined performance of each of the monolingual and bilingual groups across the three cognitive status levels. The D-KEFS composite scores were analyzed using separate ANOVAs for monolinguals and bilinguals with diagnosis as the grouping variable. The data are plotted in Figure 2 and the omnibus comparisons can be found in Table 3.

Within each language group, the main effect of diagnosis was significant for each measure, with performance decreasing with each cognitive status level. Importantly, however, the contribution of diagnostic status to performance outcomes was different for the two language groups. For monolinguals, diagnostic category explained an average of ~51% of the variance in performance, but for bilinguals it explained an average of only ~36% of the variance in group performance (see Table 3, estimates are derived by averaging within group η_p^2 values). This difference of ~14% suggests these neuropsychological tests do not fully capture bilingual performance. Across each of the three tasks, performance by monolinguals was best described by a negative linear trend with advancing dementia status. This pattern is particularly evident for the verbal fluency and Stroop tasks, particularly in their sensitivity to the difference between HA and MCI. Bilinguals follow a different trend for all three tests where performance changes were nearly quadratic (see Figure 2). Post-hoc two-way *p* values are shown on Figure 2. Age was not a significant predictor in any of the analyses. Although the expected difference between MCI and AD patients was observed, the tests did not distinguish between HA and MCI performance. Therefore, using a verbal task with bilingual individuals yields a biased score. Monolinguals have access to semantic resources and verbal cues that bilinguals do not. A fairer comparison is thus to examine within-group change scores. This was accomplished by norming performance to the HA within language group and showing that the change scores indicated greater decline for monolinguals than for bilinguals. Moreover, unlike the linear decline found for monolinguals, the decline for bilinguals across cognitive status groups was quadratic. Notably, the trail-making task, which is less verbal, also showed a quadratic trend for monolinguals.

Discussion

The present study investigated the performance of monolingual and bilingual older adults at three cognitive status levels – HA, MCI, and AD – on standard neuropsychological measures commonly used in clinical assessment. The results were examined in two ways: first, cognitive status was held constant to compare language groups within each cognitive level; second, language group was held constant to compare performance for that group across the three cognitive levels. There were two main results. First, performance of the bilinguals in the HA group was significantly lower than was found for HA monolinguals on two of the three tests. Second, comparing performance across cognitive status level, the decline over the three categories was linear for monolinguals but largely quadratic for bilinguals.

Consider first the results for the two language groups within cognitive status. For healthy older adults, there were no significant differences between monolinguals and bilinguals in age, education, or MMSE, but the monolinguals performed better than bilinguals on the Stroop and verbal fluency tasks, the two tasks that had substantial verbal components. The finding for poorer performance by bilinguals on verbal tasks is consistent with previous research (Bialystok & Luk, 2012; Gollan et al., 2007; Ivanova & Costa, 2008) and the particular pattern found for the verbal fluency task is similar as well to that reported in previous literature (Luo, Luk, & Bialystok, 2010; Roselli et al., 2002). In general it has been found that letter fluency is equivalent for monolinguals and bilinguals (but see Bialystok, Craik, & Luk, 2008b), but category fluency is consistently better for monolinguals. The present results replicate this finding.

The results for the color-word interference (Stroop) task were similar to those previously reported by Roselli et al. (2002), who showed that Spanish-English bilinguals were slower overall than their monolingual peers, particularly for the color naming condition. However, Roselli et al. presented the Stroop task in both languages, Spanish and English, and found that performance was better in the participant's native language, an effect that was tempered by language experience. Balanced bilinguals also tended to experience less interference than novices. The present results replicate the pattern of bilingual slowing on the color naming condition but equal performance with monolinguals on the word reading task as described by Roselli et al. (2002). In the present study, it was not possible to administer this task in the participant's non-English language so it is uncertain how much of the outcome score reflected language proficiency rather than cognitive level.

To summarize, this study compared performance on three neuropsychological tests from the D-KEFS battery for monolinguals and bilinguals at three levels of cognitive status. For the HA group, participants were equivalent on age, cognitive status level, and MMSE, and all participants obtained D-KEFS scores within the normal range. Nonetheless, monolinguals outperformed bilinguals on the verbal fluency and Stroop tasks. This difference in the relationship between the tests of the D-KEFS is an important indicator of bilingualism, but one that is not considered in clinical interpretations. The usual expectation would be for equivalent performance across the tests such that deviation in a particular test is interpreted as evidence of clinical impairment. In this case, however, verbal ability must be accounted for before any clinical interpretations can be made.

In the patient populations, the D-KEFS measures were largely comparable between language groups except for two scores in the AD group, supporting the notion of a standard diagnostic criterion. This unilateral approach is fine if the underlying groups do not differ. However, if systematic differences through common experience make groups diverge in these ways, the application of a single diagnostic criterion may bias outcomes. The observed differences in the AD group are in line with expected bilingual disadvantages in language production (reduced letter fluency) but advantages in cognitive control (reduction in inhibition/switching errors).

When language group was held constant, the change in scores across the three cognitive status levels was different for the monolinguals and bilinguals. For monolinguals, on average, the decline was linear, as would be expected on the basis of standardized test scores. Such a linear change across cognitive status is precisely what these tests were established to assess. This was not the case, however, for the bilingual samples. This group showed less reduction in scores from HA to MCI than their monolingual peers, and the overall pattern was nearly quadratic.

Since there is no reason to believe that the bilinguals in the random sample of healthy older adults are dementing at a faster rate than monolinguals, the present results raise an intriguing prospect for the cognitive reserve hypothesis. If cognitive decline were occurring more rapidly for bilinguals than monolinguals, then one would expect a steeper slope across the cognitive status levels for bilinguals reflecting this faster decline. This is not the case, and in fact the opposite is evident between HA and MCI. Another way of considering this point is to interpret the HA scores within each group as baseline performance levels for that language group (see Figure 2B). By standardizing performance relative to HAs with the same background within each group, the pattern of decline appears different for monolinguals and bilinguals. With each increment in dementia severity, monolingual performance drops by ~2 SD, but bilingual performance drops by ~0.5 SD. Therefore, by the time an AD diagnosis has been reached, monolinguals are ~3.5 SD from baseline but bilinguals are ~ 2.5 SDs from baseline. These data raise the possibility that many bilinguals may be misclassified as having worse cognitive status than is actually the case on the basis of inappropriate standard scores. An individual who scores only 0.5 SD from the healthy older adult mean for that group is not significantly different from that population. This finding leads to a rather surprising question: Do the bilinguals in our sample actually have MCI? Based on standard neuropsychological evaluation, a diagnosis of MCI would seem appropriate. However, once verbal fluency is accounted for, this conclusion no longer appears to be the case. Many of the MCI individuals in the present sample do not differ from healthy aging, or show only moderate decline. Nonetheless, by the time a diagnosis of AD is determined, there is no question that the individuals are impaired, although the degree of impairment may be less severe than thought. Bilinguals with a diagnosis of AD scoring ~2.5 SD below the HA mean may in fact only have declined as much as a monolingual with a diagnosis of MCI. The present results underscore the need for sensitive, non-verbal alternatives to standard neuropsychological assessment.

The portrait of bilingual aging that emerges is thus one of decelerated decline relative to monolinguals in spite of lower baseline performance, most likely because of features of the standardized tests. This finding is consistent with the cognitive reserve hypothesis. It is clear that bilingualism affects the closely intermeshed interactions of age, cognitive status (as determined by clinical assessment), and cognitive level (as measured by neuropsychological instruments). The present pattern of results likely emerges from two primary factors. First, bilinguals are known to perform more poorly on neuropsychological tests normed on English monolingual speakers. Monolinguals may be able to draw upon semantic architectures and lifetime knowledge

– resources that bilinguals would be able to access to a much greater degree in their native language. Bilinguals may also have had to translate materials in parallel to performing the task, adding cognitive load to the task. The second factor is that despite a lower initial starting point, bilinguals decline *less* than monolinguals. Thus, neuropsychological assessment and clinical evaluation should proceed not from a monolingual baseline, but rather by taking each healthy group's baseline and assessing change from that point.

Cognitive reserve is of course more complex than any single factor can fully capture. Socioeconomic status, lifestyle, overall health, education, diet and exercise are among the many factors that contribute to reserve in the face of aging. A fuller exploration of how these factors operate and interact in concert is warranted, but is beyond the scope of the present study. To accommodate this issue, the monolingual and bilingual participants in the present study were matched on as many factors as possible (including age, education and MMSE scores) to elicit an accurate measure of the effect of bilingualism on cognitive reserve.

Establishing bilingualism as a source of cognitive reserve shifts expectations for cognitive performance in this group. Healthcare providers should account for the differences in verbal ability in bilingual populations because failure to do so may result in misclassification of healthy aging as dementia. A further prediction of the present dataset, based on previous work by Schweizer et al. (2012) is that some of the bilingual individuals in the healthy sample may have greater brain atrophy than the healthy monolinguals. That is, a direct comparison might show that despite controlling for age and performance on the MMSE, some of the bilinguals may have greater frontal and temporal atrophy than expected relative to monolinguals. This hypothesis speaks directly to the *cognitive* reserve hypothesis, where experience such as education, exercise, training, or bilingualism can yield protective effects in the face of neural atrophy (Stern, 2012;

Schweizer et al., 2012). Increasing the sample size, currently in progress, will help to address this possibility.

There are two conclusions from the present study. First, neuropsychological tests may produce biased results when used with bilingual participants. This is likely due to the verbal nature of the materials. Second, bilingual performance declines less than monolingual peers across diagnostic categories. Together, these findings suggest that the D-KEFS battery is insufficiently sensitive to bilingual baseline performance and may mask cognitive decline when this is ignored, especially at the earliest signs of decline. The finding that the neuropsychological scores surveyed in the present study did not capture as much variance in bilingual performance as it did for monolinguals suggests a need for more sensitive tests or at least more differentiated scoring criteria that can account for these experiences. This point is of particular concern since cognitive reserve can stave off apparent decline despite comparable levels of neurodegeneration. Identifying this earlier could lead to more accurate diagnostic outcomes for bilinguals.

References

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., . . . Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology, 81,* 1938-1944.

Alladi, S., Bak, T. H., Mekala, S., Rajan, A., Chaudhuri, J. R., Mioshi, E., ... & Kaul, S. (2016). Impact of bilingualism on cognitive outcome after stroke. *Stroke, 47,* 258-261. doi: 10.1161/STROKEAHA.115.010418

Bak, T. H. (2016). The impact of bilingualism on cognitive aging and dementia: Finding a path through a forest of confounding variables. *Linguistic Approaches to Bilingualism, 6,* 205-226. doi:10.1075/lab.15002.bak

Bak, T. H., & Alladi, S. (2014). Can being bilingual affect the onset of dementia? *Future Neurology, 9,* 101-103.

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics, 35*, 857-894. doi: 10.1017/S0142716414000174

Bedore, L. M., & Peña, E. D. (2008). Assessment of Bilingual Children for Identification of Language Impairment: Current Findings and Implications for Practice. *International Journal of Bilingual Education and Bilingualism, 11,* 1-29. doi: 10.2167/beb392.0

Bennett, D. A., Schneider, J. A., Tang, Y., Arnold, S. E., & Wilson, R. S. (2006). The effect of social network on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurology, 5,* 406-412.

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Mendes de Leon, C. F., Arnold, S. E., … Bienias, J. L. (2003). Education modifies the relation of AD pathology to level of cognitive function in older persons. *Neurology, 60,* 1909 – 1915.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12,* 3-11. doi: 10.1017/s1366728908003477

Bialystok, E. (in press). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism*. doi:10.1075/lab.15040.bia

Bialystok, E., Craik, F. I., Binns, M. A., Ossher, L., & Freedman, M. (2014). Effects of bilingualism on the age of onset and progression of MCI and AD: Evidence from executive function tests. *Neuropsychology*, 28, 290-304.

Bialystok, E., Craik, F. I., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24,* 40-49.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19,* 290-303. doi: 10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2008a). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 859-873.

Bialystok, E., Craik, F. I. M., & Luk, G. (2008b). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics, 21,* 522-538.

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition, 15,* 397-401

Cattell, R. B., & Cattell, A. K. S. (1973). *Measuring intelligence with the culture fair tests*. Champaign, IL: Institute for Personality and Ability Testing.

Chaoimh, D. N., De Bhaldraithe, S., O'Malley, G., Bhuí, C. M. A., & O'Keeffe, S. T. (2015). Importance of different language versions of cognitive screening tests: Comparison of Irish and English versions of the MMSE in bilingual Irish patients. *European Geriatric Medicine*, *6*(6), 551-553.

Craik, F. I. M., Bialystok, E., & Freedman, M. (2010). Delaying the onset of Alzheimer disease Bilingualism as a form of cognitive reserve. *Neurology, 75,* 1726-1729. doi: 10.1212/WNL.0b013e3181fc2a1c

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system (D-KEFS)*. San Antonio, TX: Psychological Corporation.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12(3)*, 189-198.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience, 33*, 387-396. doi: 10.1523/JNEUROSCI.3837-12.2013

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston naming test performance. *Journal of the International Neuropsychological Society, 13,* 197–208.

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., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16*(4), 562-576.

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

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology, 26*, 640-648.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: preliminary norms in Spanish and English. *Brain and language, 65*(3), 422-440.

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a "bilingual advantage" in stroop interference in a nonimmigrant sample. *Quarterly Journal of Experimental Psychology*, 65(2), 356-369.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497-514.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142(2)*, 573-603.

Lancu, I., & Olmer, A. (2006). The minimental status examination – an up-to-date review. *Harefuah*, 145(9), 687-690.

Landau, S. M., Mintun, M. A., Joshi, A. D., Koeppe, R. A., Petersen, R. C., Aisen, P. S., … Jagust, W. J. (2012). Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of Neurology, 72,* 578-586.

Lopez, O. L., Becker, J. T., Wahed, A. S., Saxton, J., Sweet, R. A., Wolk, D. A., … Dekosky, S. T. (2009). Long-term effects of the concomitant use of memantine with cholinesterase inhibition in Alzheimer disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 80,* 600–607.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605-621.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114,* 29-41.

Massoud, F., & Gauthier, S. (2010). Update on the pharmacological treatment of Alzheimer's disease. *Current Neuropharmacology, 8*, 69-80.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232-258. doi:10.1016/j.cogpsych.2012.12.002

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French–English bilingual children with SLI: How do they compare with their monolingual peers? *Journal of Speech, Language, and Hearing Research, 46,* 113-127.

Peña, E. D., Gutierrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *BESA: Bilingual English-Spanish Assessment Manual.* AR-Clinical publications.

Rivera-Mindt, M., Arentoft, A., Germano, K. K., D'Aquila, E., Scheiner, D., Pizzirusso, M., … Gollan, T. H. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review*, *18*(3), 255–268. http://doi.org/10.1007/s11065-008-9069-7.

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.

Rockwood, K. (2004). Size of the treatment effect on cognition of cholinesterase inhibition in Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 75,* 677-685. doi: 10.1136/jnnp.2003.029074

Roselli, M., Ardila, A., Santisi, M. N., Arecco Mdel, R., Salvatierra, J., Conde, A., & Lenis, B. (2002). Stroop effect in Spanish-English bilinguals. *Journal of the International Neuropsychological Society*, 8(6), 819-827.

Schweizer, T. A., Ware, J., Fischer, C. E., Craik, F. I. M., & Bialystok, E. (2012). Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex, 48(8),* 991-996.

Starr, J. M. (2010). Cognitive impairment in older adults: a guide to assessment. *Clinical Medicine*, *10*(6), 579-581.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8,* 448-460.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, *11*(11), 1006-1012.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Journal of American Medical Association, 271,* 1004-1010.

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

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental status examination: A comprehensive review. *Journal of American Geriatrics Society*, 40(9), 922-935.

Valenzuela, M. J., Sachdev., P., Wen, W., Chen, X., & Brodaty, H. (2008). Lifespan mental activity predicts diminished rate of hippocampal atrophy. *PLoS ONE. 3,* e2598.

Wodniecka, Z., Craik, F. I. M., Luo, L., & Bialystok, E. (2010). Does bilingualism help memory? Competing effects of verbal ability and executive control. *International*

Journal of Bilingual Education and Bilingualism, 13, 575–595.

Zhu, C. W., Livote, E. E., Scarmeas, N., Albert, M., Brandt, J., Blacker, D., . . . Stern, Y. (2013). Long-term associations between cholinesterase inhibitors and memantine use and health outcomes among patients with Alzheimer's disease. *Alzheimers & Dementia, 9,* 63-75.

Table 1. Descriptive statistics for background measures and executive function tasks for the three cognitive status groups

	Healthy Older Adults		Mild Cognitive Impairment		Alzheimer's Disease	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
			Demographic Information			
Age at Testing	74.9 (4.6)	74.7 (3.9)	66.5 (12.3)	70.0 (10.7)	74.2 (11.2)	81.4 (8.4)
Education in Years	17.5 (4.0)	16.7 (2.7)	15.5 (3.8)	14.3 (3.9)	12.5 (3.7)	12.2 (4.9)
Onset Age			62.8 (13.4)	66.2 (11.7)	72.1 (10.4)	78.8 (8.9)
MMSE	29.1(0.9)	29.0 (1.5)	29.0 (1.4)	28.6 (1.8)	22.3 (4.9)	20.8 (5.6)
Standardized Scores			Trail Making Task			
Number sequencing	12.6 (2.4)	10.9 (3.6)	9.5 (3.5)	9.0 (4.6)	4.1 (3.8)	4.1 (3.8)
Switching	12.0 (3.1)	10.1 (3.4)	9.9 (2.9)	8.4 (3.6)	4.8 (4.1)	5.4 (3.4)
Switching errors	11.8 (0.4)	10.9 (1.7)	10.9 (1.7)	10.0 (3.2)	7.2 (4.5)	8.8 (3.4)
			Verbal Fluency			
Letter fluency	13.9 (3.1)	12.4 (3.8)	11.3 (3.6)	10.8 (3.6)	7.8 (4.3)	5.7 (3.8)
Category fluency	13.2 (3.6)	9.8 (4.2)	8.9 (3.0)	8.0 (3.0)	5.4 (3.1)	4.7 (3.4)
Category switching (total correct)	13.6 (2.9)	10.1 (4.0)	9.7 (3.5)	9.2 (2.9)	5.1 (3.5)	4.8 (3.1)
Category switching (total switching accuracy)	13.4 (2.5)	10.5 (3.6)	9.2 (3.3)	9.0 (2.5)	4.8 (3.5)	4.9 (3.1)
Percent set-loss errors	11.7 (1.5)	11.1 (2.8)	10.2 (3.5)	10.4 (3.1)	7.0 (4.4)	6.5 (4.6)
Percent repetition errors	11.4 (1.4)	9.4 (2.8)	9.9 (2.8)	10.1 (2.8)	7.4 (4.5)	8.8 (3.7)
percent switching accuracy	12.9 (0.2)	12.8 (0.6)	9.9 (3.0)	10.8 (2.2)	7.3 (3.6)	8.1 (3.9)
			Stroop Task			
Color naming	11.4 (1.8)	9.2 (2.8)	9.1 (3.1)	8.9 (2.8)	5.6 (4.1)	4.8 (4.1)
Word reading	10.9 (2.3)	10.8 (2.5)	9.4 (3.3)	9.2 (3.2)	8.0 (3.9)	7.2 (4.4)
Inhibition (Stroop effect)	12.6 (1.9)	10.8 (2.3)	9.5 (3.5)	9.5 (2.9)	6.2 (4.7)	6.3 (4.2)
Inhibition/switching	12.8 (1.2)	10.4 (2.6)	8.6 (3.4)	8.8 (2.9)	5.0 (4.3)	5.2 (4.1)
Inhibition/Switching vs. Inhibition	10.3 (2.1)	9.6 (3.6)	10.2 (2.9)	10.5 (2.6)	9.9 (3.4)	11.5 (3.8)
Inhibition errors	12.3 (1.1)	10.9 (2.6)	9.2 (3.7)	9.7 (3.4)	8.6 (4.4)	8.0 (4.0)
Inhibition/Switching errors	11.7 (1.8)	10.4 (2.9)	8.8 (3.2)	8.7 (3.5)	2.8 (3.5)	5.8 (4.1)

Table 2. Bayesian comparisons of monolingual and bilingual participants for all measures at each cognitive status.

	Bayesian t-tests								
	Healthy Older Adults			Mild Cognitive Impairment			Alzheimer's		
	Difference of the means	95% Credible Interval on the mean difference	% Mean Difference Posterior > 0	Difference of the means	95% Credible Interval on the mean difference	% Mean Difference Posterior > 0	Difference of the means	95% Credible Interval on the mean difference	% Mean Difference Posterior > 0
	Demographic Information								
Age at Testing	0.09	[-3, 3.3]	52.3%						
Education	0.79	[-1.79, 3.2]	74.4%						
MMSE	-0.06	[-0.83, 0.73]	43.3%	0.39	[-0.25, 0.99]	89.1%	1.50	[-0.61, 3.6]	91.6%
<u>*Standardized Scores*</u>	Trail Making Task								
Number sequencing	1.50	[-0.75, 3.6]	91.5%	0.58	[-1.4, 2.6]	72.0%	0.06	[-1.7, 1.9]	52.5%
Number-Letter Switching	2.10	[-0.17, 4.5]	96.1%	1.60	[-0.035, 3.2]	97.3%	-0.62	[-2.8, 1.6]	28.7%
Number Letter Switching errors	0.00	[-0.029, 0.026]	50.8%	0.15	[-0.097, 0.63]	88.4%	-1.60	[-4.1, 0.79]	8.8%
	Verbal Fluency Task								
Letter fluency	1.50	[-1, 4.1]	87.6%	0.44	[-1.2, 2.2]	69.7%	2.10	[0.059, 4.2]	97.6%
Category fluency	3.40	[0.44, 6.2]	98.9%	0.85	[-0.56, 2.3]	88.0%	1.00	[-0.68, 2.6]	88.5%
Category switching (total correct)	3.60	[0.95, 6.2]	99.6%	0.50	[-1.1, 2.1]	73.6%	0.25	[-1.4, 2.0]	61.5%
Category switching (total accuracy)	3.00	[0.68, 5.3]	99.3%	0.23	[-1.2, 1.6]	62.6%	-0.18	[-1.8, 1.4]	41.5%
Percent set-loss errors	0.06	[-1.1, 1.5]	53.9%	-0.03	[-1.5, 1.4]	48.2%	0.52	[-1.8, 2.8]	67.8%
Percent repetition errors	2.00	[0.32, 3.6]	99.1%	-0.14	[-1.5, 1.2]	41.8%	-1.40	[-3.6, 0.72]	10.0%
Percent switching accuracy	0.00	[-0.00026, 0.00025]	50.1%	-0.46	[-1.7, 0.67]	21.2%	-0.91	[-2.8, 1.1]	17.5%
	Stroop Task								
Color naming	2.30	[0.53, 4.1]	99.3%	0.24	[-1.2, 1.6]	63.3%	0.91	[-1.3, 3.1]	79.2%
Word reading	0.27	[-1.5, 2.1]	62.1%	0.12	[-1.5, 1.7]	56.1%	0.90	[-1.4, 3.2]	78.5%
Inhibition	1.90	[0.32, 3.4]	99.1%	-0.06	[-1.7, 1.6]	47.1%	-0.10	[-2.8, 2.6]	47.0%
Inhibition/Switching	2.40	[0.76, 3.9]	99.8%	-0.21	[-1.8, 1.3]	39.3%	-0.15	[-3.1, 2.8]	45.8%
Inhibition/Switching vs. Inhibition	0.46	[-1.7, 2.6]	67.4%	-0.36	[-1.7, 0.97]	29.6%	-1.50	[-3.5, 0.48]	7.1%
Inhibition errors	1.00	[0.24, 1.9]	99.7%	-0.47	[-2.2, 1.3]	29.8%	0.60	[-1.9, 3.0]	68.4%
Inhibition/Switching errors	1.10	[-0.41, 2.6]	92.9%	0.09	[-1.6, 1.8]	54.2%	-4.70	[-7, -2.5]	0.1%

Larger mean differences are consistent with higher values for monolinguals relative to bilinguals and vice-versa. The percentage values reflect the proportion of the posterior difference parameters that exceeded 0.

Table 3. Omnibus effects across cognitive status groups for composite scores of D-KEFS measures

	df	*F*	*p*	η_p^2
Monolinguals				
Stroop composite	2, 80	27.35	0.0000	0.41
Verbal fluency composite	2, 85	53.39	0.0000	0.56
Trail making test composite	2, 84	51.68	0.0000	0.55
Bilinguals				
Stroop composite	2, 81	18.93	0.0000	0.32
Verbal fluency composite	2, 83	26.28	0.0000	0.39
Trail making test composite	2, 87	26.58	0.0000	0.38

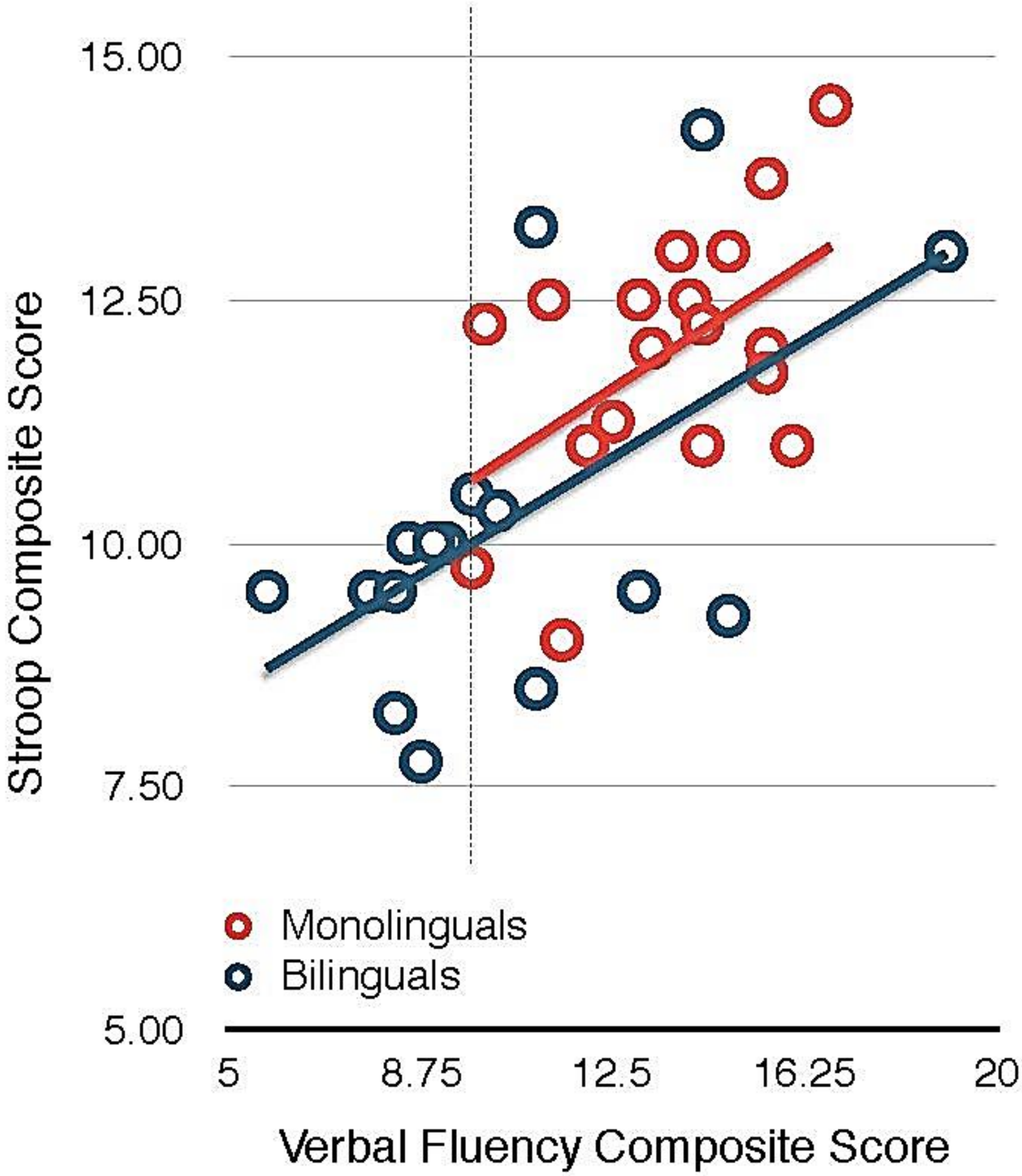
15.00
12.50
10.00
7.50
5.00
Stroop Composite Score
Monolinguals
Bilinguals
5
8.75
12.5
16.25
20
Verbal Fluency Composite Score

Figure 1.

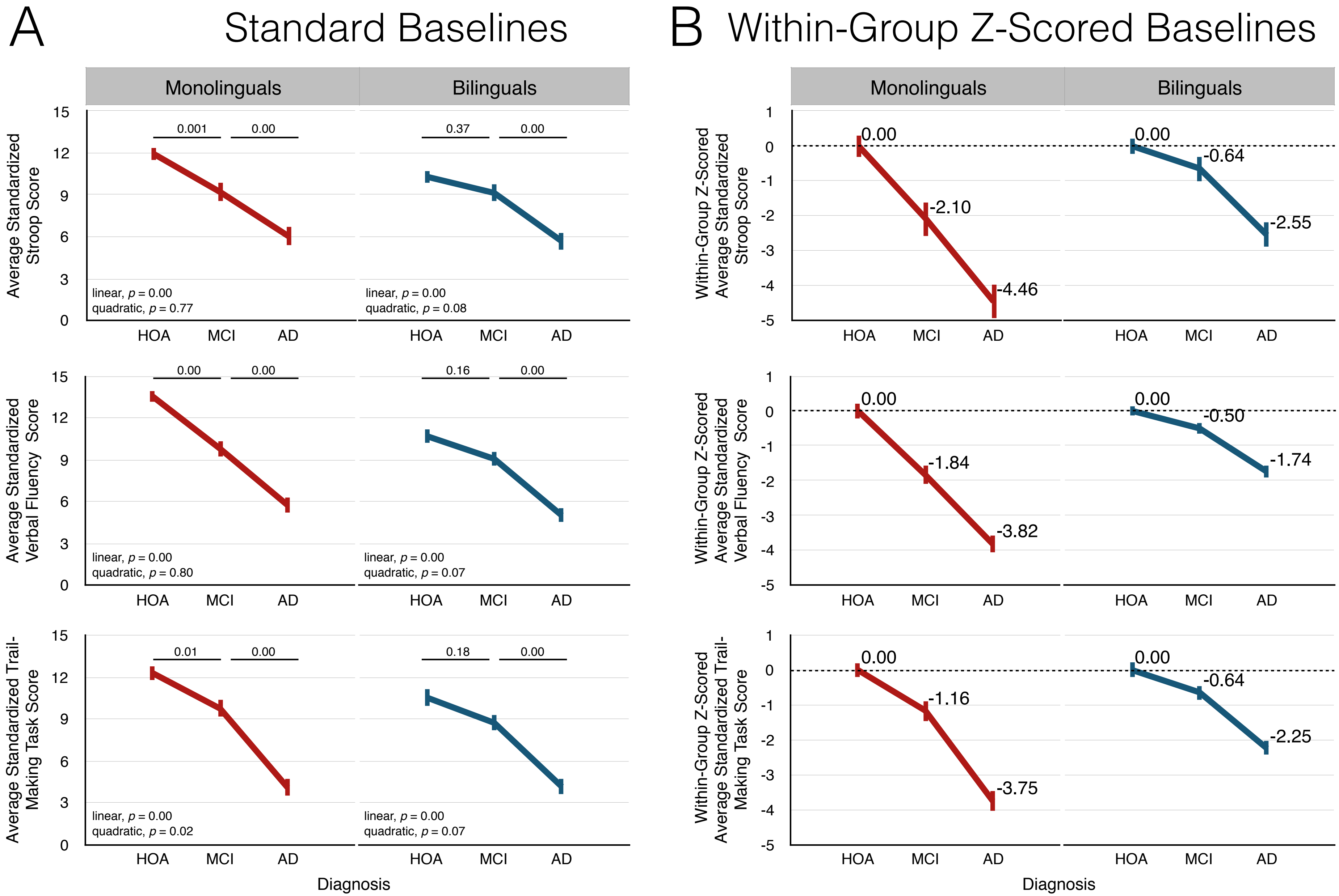
A
Standard Baselines
B
Within-Group Z-Scored Baselines
Monolinguals
Bilinguals
Average Standardized Stroop Score
Average Standardized Verbal Fluency Score
Average Standardized Trail-Making Task Score
Within-Group Z-Scored Average Standardized Stroop Score
Within-Group Z-Scored Average Standardized Verbal Fluency Score
Within-Group Z-Scored Average Standardized Trail-Making Task Score
HOA
MCI
AD
0.001
0.00
0.37
0.16
0.01
0.18
linear, p = 0.00
quadratic, p = 0.77
quadratic, p = 0.08
quadratic, p = 0.80
quadratic, p = 0.07
quadratic, p = 0.02
0.00
-2.10
-4.46
-0.64
-2.55
-1.84
-3.82
-0.50
-1.74
-1.16
-3.75
-2.25
Diagnosis

Figure 2.

Figure Captions

Figure 2. Healthy older adult composite verbal fluency scores predicting composite Stroop scores. Data are divided by language group. The dashed line indicates the lowest monolingual verbal fluency scores.

Figure 2. Average standardized performance on three measures from the D-KEFS by group. Error bars are +/- 1 SEM. Post-hoc Tukey *p*-values appear for incremental comparisons (i.e. HOA – MCI and MCI – AD). For each comparison, HOA were different from AD, $p = 0.000$. Panel A represents norm-referenced scores, panel B represents scores standardized by within-group healthy older adult performance.