

Preliminary Data on the Effect of Culture on the Assessment of Alzheimer's Disease-Related Verbal Memory Impairment with the International Shopping List Test

Yen Ying Lim^{1,*}, Robert H. Pietrzak², Peter J. Snyder³, David Darby^{4,5}, Paul Maruff^{4,5}

¹Department of Psychiatry, University of Melbourne, Parkville, Australia

²Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

³Lifespan Hospital System and Department of Neurology, The Warren Alpert Medical School of Brown University, Providence, RI, USA

⁴CogState Ltd, Melbourne, Australia

⁵Centre for Neuroscience, University of Melbourne, Parkville, Australia

*Corresponding author at: Department of Psychiatry, University of Melbourne, Level 2, 255 Bourke Street, Melbourne, VIC 3000, Australia.

Tel: +61-3-9388-1633; fax: +61-3-9387-5061.

E-mail address: limyy@student.unimelb.edu.au (Y.Y. Lim).

Accepted 3 November 2011

Abstract

The International Shopping List Test (ISLT) is a measure of verbal learning and memory, developed specifically for use in people from different cultural and linguistic backgrounds. In this report, we describe two studies that examined the ISLT's ability to detect memory impairment and memory decline in patients with mild Alzheimer's disease (AD) from a range of cultural and linguistic backgrounds. In Study 1, the performance of Australian-English-speaking adults with mild AD was compared with that of native Australian-English- and Korean-speaking patients with mild AD. Compared with controls, patients with AD from both language groups showed large but equivalent impairments in total recall, delayed recall, rate of learning, and primacy and retention-weighted recall (RWR) measures on the ISLT. In Study 2, the rate of deterioration in verbal memory over 1 year was examined in groups of native Canadian-English, French, and Korean speakers with mild AD using the total recall, delayed recall, and RWR measures. Rates of change on all three measures were equivalent across the language groups, although the magnitude of deterioration was most pronounced for the total recall and RWR measures. Taken together, these results suggest that the ISLT is valid and reliable for the assessment of verbal learning and memory impairment and decline in patients with mild AD from diverse language groups.

Keywords: Alzheimer's disease; Cross-cultural/minority; Learning and memory

Introduction

The assessment of verbal episodic memory is central to the neuropsychology of Alzheimer's disease (AD) and is conducted most commonly using verbal list learning tests (Buschke et al., 2006; Mitrushina, Boone, Razani, & D'elia, 2005). In mild AD, impairment on verbal list learning tests manifests as a reduced number of total words recalled, slower rates of learning across trials, decreased ability to recall words after a delayed period of time, and reduced primacy with relatively normal recency (Bayley et al., 2000; Greenway et al., 2006; Jungwirth et al., 2009). In addition, indices such as retention-weighted recall (RWR), derived from the combination of primacy and total recall scores, also show large impairments in AD (Buschke et al., 2006).

Given that AD presents in older people from all countries, it would be useful to have verbal list learning tests that can be used in different cultural or language groups with equivalent validity and reliability. However, because verbal list learning tests have been developed in Western Europe and North America, their application to cultures where English or other modern European languages are not spoken has been limited (Ardila, 2005; Greenfield, 1997). One method that has been used to

adapt standardized list learning tests to other cultures has been to directly translate the original English words into the language of interest (Reynolds, 2000). For example, the memory test from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery has been translated from American English into Korean (Lee et al., 2002) and French (Demers et al., 1994) through a process of translation and back-translation. However, this approach has been criticized because certain languages may lack the original words or direct translation may not be possible (Dick, Dick-Muehlke, & Teng, 2006). For example, the English word "ranger" used on the RAVLT has no direct equivalent in Mandarin, with 森林管理员 (person in charge of forest) providing the closest translation. Obviously, this phrase is different to the original single word "ranger," which can refer to a range of occupations and is not strictly restricted to the protection of forests. Furthermore, the familiarity and relevance of the term for a person in charge of a forest to Mandarin-speaking people may be low. Therefore, despite translation, the two words are not equivalent. Cultural relevance can also affect memory for English words. Recently, we found that memory performance in healthy English-speaking adults living in Australia was almost 1 *SD* below that of a matched English-speaking sample from the USA when the stimuli on the verbal list learning test consisted of nouns describing food items more common in the USA than in Australia (e.g., bagel, soda; Lim et al., 2009). This finding persisted despite the USA food item stimuli being relatively well known to the Australian group. However, this culture-based asymmetry was reversed completely when the verbal list learning test was modified so that stimuli were food items more common in Australia than in the USA (Lim et al., 2009).

A second approach to adapting verbal list tests to other cultures has been to translate words but ensure that the translated and original words match according to some lexical characteristics within each language. For example, list learning stimuli have been matched on characteristics known to influence encoding, such as frequency, familiarity, imaginability, syllable number, or length (Bock & Klinger, 1986; Glanzer, 1972; Paivio, 1968). This approach requires reliable information about the different linguistic characteristics for the language of interest. However, for many languages, this information does not exist and even when available, language use can change rapidly over time, often rendering it inaccurate (Ardila, 2007). Furthermore, in Asian languages (e.g., Korean), where pictorial characters are used instead of an alphabet (e.g., the Roman alphabet), characteristics such as syllable number or word length cannot be used to translate words. For example, the fruit "apple" in English contains two syllables, five letters, and five pen-strokes, whereas "apple" in Korean (사과) consists of two syllables, two characters, and nine pen-strokes. Even with the use of "pīn yīn," the Romanization of Chinese characters to aid pronunciation ("apple" written as "píng guǒ"), there is the added complexity of the four separate tones in Mandarin that must be considered. Thus, matching by word characteristics is limited in the extent to which it can be applied universally.

Individuals from some cultures may also have lower levels of test-taking experience (Ardila, 2005). Consequently, assumptions about values, knowledge, and communication, which are implicit in verbal memory measures, are often not met when these assessments are conducted in other cultures (Greenfield, 1997). This can consequently reduce the validity of performance scores derived from those tests (Reynolds, 2000; Shiraev & Levy, 2007). For example, in North America and Western Europe, testing in general is quite common, and the dominant mode of such testing involves asking the examinee for information that the examiner does not necessarily lack (Greenfield, 1997). This type of information gathering may be unfamiliar to people from cultures where questions are asked only to gather information that is lacking (Greenfield, 1997). Lower familiarity with test taking has also been recognized to reduce test validity in subpopulations from developed countries, for example, in older adults with low educational experience, people from rural or remote regions, people from low socioeconomic groups, or immigrants (Ardila, 2005; Greenfield, 1997; Nell, 2000; Youngjohn, Larrabee, & Crook, 1991). One solution to the issue of decreased familiarity with test-taking has been to develop neuropsychological tests that approximate situations from everyday life settings (e.g., Wilson, Cockburn, Baddeley, & Hiorns, 1989). For verbal list learning tests, one attempt at such approximation has been to include items commonly used in everyday situations in the list (e.g., the California Verbal Learning Test, CVLT; Delis, Kramer, Kaplan, & Ober, 1987; and the Grocery List Selective Reminding Test; Youngjohn et al., 1991). Importantly, although these tests contain items that are found commonly in Western societies, they do not necessarily possess cultural appropriateness when adapted for use in other cultures or languages (e.g., items on the CVLT such as slacks and chisel).

Recently, we developed a method for the cross-cultural adaptation of a verbal list learning test. This method is based on the shopping list model, with the equivalence of stimulus words across languages or cultures controlled by restricting the selection of word stimuli to items that are common in the shops, stores, or markets (hereafter termed stores) of specific geographic locations (Lim et al., 2009). The International Shopping List Test (ISLT) is a four-trial (three learning trials and one delayed recall trial) verbal list learning test in which individuals are instructed to remember a list of 12 items that they need to obtain from their local store. To ensure that items on the shopping list are relevant to the region of interest, a large pool of common shopping list items (128 items) is translated into the language of interest by a certified translator from that same region. After translation, a web-based survey is conducted in which 30 or more people from the region of interest are asked to rate each shopping list item according to the ease (i.e., "very difficult," "difficult," "easy," or "very easy") with which it can be obtained locally.

Only items rated as “very easy” or “easy” to obtain are included as stimuli on the ISLT (Lim et al., 2009). This procedure typically yields 96 words, which are organized into eight different ISLT lists of 12 words for each language (see Appendix for Australian-English lists). In an earlier study conducted on healthy young adults from Mandarin, Malay, French, and English language groups, who were matched on age, education, and gender, performance on the ISLT was equivalent on measures of total recall, primacy, recency, and rate of recall (Lim et al., 2009). Two recent studies evaluated the ability of the ISLT to detect memory impairment in clinical groups. In the first study, the magnitude of impairment in English-speaking people with schizophrenia was estimated to be equivalent by the Hopkins Verbal Learning Test (HVLT) and ISLT, and there was a high correlation between performance on the two memory tests (Pietrzak et al., 2009). A second study observed large and reliable impairment in performance on the ISLT in English-speaking adults with mild AD compared with matched controls (Thompson et al., 2011). These studies provide a sound basis for evaluating the ability of the ISLT to detect AD-related memory impairment in languages other than English.

Reynolds (2000) has recommended that evaluating a test’s ability to detect impairment in a criterion disorder in different cultures can provide a good challenge of the cultural equivalence of the test. In this context, memory impairment in AD provides a good criterion, as the magnitude of such impairment is large and well described (Baddeley, 1992; Perry, Watson, & Hodges, 2000). The disease can also be diagnosed clinically, and its severity staged with high reliability in diverse cultural groups (e.g., Chow et al., 2002; Kalaria et al., 2008). Finally, the severity of AD-related memory impairment increases in magnitude over time and the extent to which a test can detect changes in memory over time in different language groups provides another opportunity for determining cultural equivalence.

Thus, the aim of Study 1 in the present report was to replicate and extend the findings of Thompson and colleagues (2011) by comparing the nature and magnitude of impairment on the ISLT in Australian-English and Korean speakers with mild AD. We hypothesized that relative to controls, both Australian and Korean patients with AD would show impaired performance on the ISLT characterized by impaired total recall scores, impaired delayed recall scores, lower RWR scores, lowered primacy effect, intact recency effect, and lower rate of learning. The first aim of Study 2 was to identify performance measures that would be useful for characterizing AD-related memory decline over a period of 1 year. The second aim of Study 2 was to determine the extent to which ISLT performance changed with disease progression in Canadian-English, Korean, and French speakers with mild AD. We hypothesized that decline in performance on the ISLT over time would not be influenced by the culture or the language of the AD group.

Study 1

Method

Participants. The control group consisted of 30 healthy Australian adults aged 60 or older who were recruited from an ongoing study of healthy cognitive aging (Frederickson et al., 2010). They were recruited via an advertizing campaign through articles in advocacy group newsletters, and a news release from the advocacy group reproduced in several local newspapers. All individuals had been undergoing serial cognitive assessment and were selected because any aspect of their cognitive function had not changed over the previous 2 years and because they did not complain of any memory problems. All had been evaluated by a board certified behavioural neurologist (D.D.) who determined that they had no evidence of cognitive, neurological, or psychiatric symptoms and that they met the inclusion/exclusion criteria as detailed below. None had been tested with the ISLT.

Thirty patients with AD were recruited from AD clinics in major hospitals in Melbourne, Australia, and Seoul, Korea. All of the Australian sample were Caucasian and all of the Korean sample Asian. Additionally, 93% of the Australian sample were born in Australia (the two non-Australian born adults were born in Italy and had lived in Australia for 53 and 55 years), but all indicated that English was their primary language. All Korean samples were born in Korea and were native Korean speakers. All patients with AD met diagnostic criteria for probable AD as defined by the National Institute of Neurologic and Communicative Disorders and Stroke–AD and Related Disorders Association (NINCDS-ADRDA; McKhann et al., 1984) and had an MRI brain scan supporting the clinical diagnosis of AD. Additional inclusion criteria were a score of 18–26 on the Mini-Mental State Examination (MMSE), a Rosen-Modified Hachinski Ischemic score of <4 , and written informed consent from the patient and the patient’s caregiver for the original test protocol. Patients were rated using the Clinical Dementia Rating (CDR) scale to provide a sum of boxes score and all were taking cholinesterase inhibitors at the time of assessment and had been doing so for 6 months or longer. Patients with AD had a CDR sum of boxes score of 1 or more and were excluded if they had any one or more of the following: a neurological disease other than AD that might affect cognition; a major psychiatric disorder, systemic illness, or symptoms that could affect the patient’s ability to complete the study; a Geriatric Depression Scale score of ≥ 10 ; or if they have used anticonvulsant, antiparkinsonian, anticoagulant, narcotic, or immunosuppressive medications within 3 months prior to assessment. No participant in the study had been tested using the ISLT

in the past. The Korean and Australian AD groups did not differ with respect to age, education, MMSE scores, or depression. Data collected from participants complied with the regulations of institutional research and ethics committees. Demographic and clinical characteristics of the control and AD groups are shown in Table 1.

Measures. Eight lists of the 12-word, 4-trial (three learning trials and one delayed recall trial) ISLT was adapted to ensure cultural and linguistic relevance to the Australian-English and Korean groups using the methodology outlined by Lim and colleagues (2009) and described in the Introduction section. The eight versions of the Australian list are shown in the Appendix. In each assessment, one of the eight lists was chosen at random by the computer software. For each list, the order in which items were presented was randomized between participants by the computer software, with the order of the items remaining constant across the three learning trials. Statistics showing the comparability of performance on the different Australian shopping lists, derived from the sample of healthy older adults described in Thompson and colleagues (2011), are also provided in the Appendix. During each assessment, the computer presented the words to the examiner one at a time at a rate of 1 word per 2 s. The participant was instructed to “try and remember as many items on the shopping list as possible.” The examiner then read each item to the participant as they appeared on the screen. The computer screen was never visible to the participant. Once all 12 words had been read to the participant, they were instructed to recall as many items from the list as possible with the statement “tell me as many items on the shopping list as you can remember.” The list of 12 words is displayed on the computer screen and as the participant recalled each item, the examiner marked the item by clicking on the relevant checkbox. If words were repeated, the checkbox was clicked again. Another checkbox was clicked if the participant said a word that was not on the original list (i.e., an intrusion). When the participant indicated that no more items could be recalled, the trial was stopped and the same process was repeated two more times. For the delayed recall trial, participants were asked to recall as many items as possible from the initial list after a delay of 15 min.

Data analysis. For each participant, for each trial, total number of words recalled correctly was computed. For correctly recalled words, the order in which they were recalled was also recorded and given an RWR score, where words recalled at the start of the list are given higher values than those recalled at the end of the list (Buschke et al., 2006). A total RWR score and total words recalled across learning trials were also recorded. A primacy score was computed for each subject by summing the number of words recalled from the first four positions on the list recalled on each trial (maximum score = 12). A recency score was also computed by summing the number of words recalled from the last four positions on the list recalled on each trial (maximum score = 12). Serial position curves were also constructed for free recall data for both healthy and AD participants (Fig. 1). Total recall, delayed recall, total RWR, primacy, and recency scores were compared using a series of one-way between-groups (control and two AD groups) analysis of variance (ANOVA). Data for words recalled and RWR at each trial were then submitted to separate 3 (group: controls, Australian AD, Korean AD) \times 3 (trial: 1, 2, 3) repeated-measures ANOVAs. The Pearson correlations were computed to examine bivariate associations between the total recall, delayed recall, and total RWR measures.

Table 1. Participant characteristics

Measure	Study 1			Study 2		
	Australian control group	Australian AD group	Korean AD group	Canadian English	French	Korean
<i>n</i>	30	30	30	20	10	20
Race	30 Caucasian	30 Caucasian	30 Asian	20 Caucasian	9 Caucasian/1 African	20 Asian
<i>n</i> born in country of assessment	30	28	30	20	10	20
Age	68.2 (8.5)	69.3 (11.9)	71.3(10.2)	71.3 (6.5)	70.3 (6.9)	69.9 (7.2)
Gender (% men)	53.3	56.6	53.2	56.6	50	54
Years of education	12.1 (2.9)	11.7 (3.2)	10.9 (3.7)	11.6 (6.7)	10.8 (7.8)	12.7 (1.2)
MMSE	29.1 (.9)	22.3 (3.3)	22.1 (5.9)	20.4 (4.5)	21.5 (3.5)	21.6 (3.4)
CDR-SOB	0	4.9 (2.3)	4.2 (3.1)	4.1 (2.1)	4.5 (3.2)	4.8 (2.6)
Depression score (GDS)	1.5 (2.0)	1.7 (2.1)	2.1 (1.9)	1.3 (2.3)	1.5 (2.3)	1.9 (2.3)

Notes: AD = Alzheimer's disease; MMSE = Mini-Mental State Examination;
CDR-SOB = Clinical Dementia Rating, Sum of Boxes;
GDS = Geriatric Depression Scale.

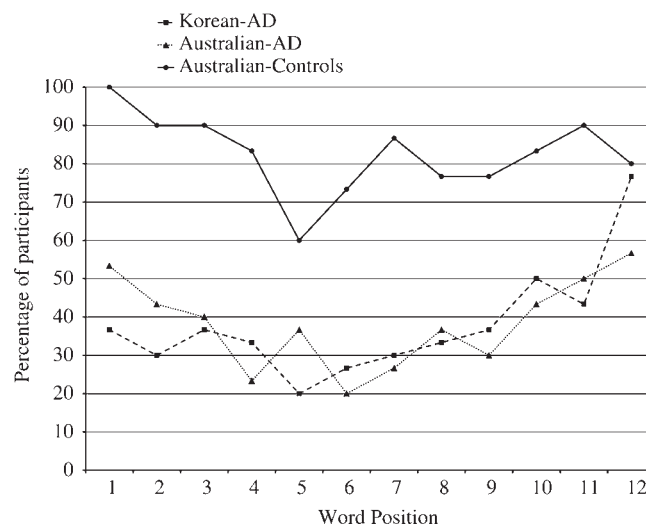


Fig. 1. Serial position curve for Korean AD, Australian AD, and Australian controls on Trial 3 of the ISLT, Study 1.

Table 2. Group mean performance on ISLT outcome measures comparing Australian control, Australian AD, and Korean AD groups, Study 1; one-way ANOVAs compared performance on total recall, delayed recall, total RWR, primacy, and recency only

Measure	Australia (English)		Korea (Korean) AD (<i>N</i> = 30)	One-way ANOVA (<i>F</i>)	<i>Post hoc</i> pairwise comparisons using Bonferroni		
	Controls (<i>N</i> = 30)	AD (<i>N</i> = 30)			Controls versus Australian AD (Cohen's <i>d</i>)	Controls versus Korean AD (Cohen's <i>d</i>)	Australian AD versus Korean AD (Cohen's <i>d</i>)
Recall							
Trial 1	6.63 (1.61)	3.00 (1.55)	2.67 (1.37)				
Trial 2	9.10 (1.52)	4.07 (1.68)	4.33 (1.58)				
Trial 3	9.90 (1.58)	4.60 (1.59)	4.80 (1.90)				
Total	25.63 (3.65)	11.67 (4.31)	11.80 (3.89)	133.96**	3.50**	3.96**	0.14
Delayed	8.67 (1.69)	1.90 (1.67)	1.50 (1.74)	166.56**	4.03**	4.14**	0.21
RWR score							
Trial 1	45.03 (10.60)	17.50 (10.45)	13.87 (9.79)				
Trial 2	61.93 (10.38)	23.80 (12.32)	23.97 (13.48)				
Trial 3	65.53 (9.47)	28.97 (13.29)	27.47 (15.47)				
Total	172.50 (23.42)	70.27 (31.56)	65.30 (32.48)	136.16**	3.68**	4.02**	0.21
Primacy							
Total	9.60 (1.48)	3.77 (2.76)	3.30 (2.64)	68.36**	2.62**	3.06**	0.06
Recency							
Total	8.53 (1.76)	4.87 (3.01)	5.77 (2.27)	20.30**	1.49**	1.56**	−0.13

Notes: Degrees of freedom for all one-way ANOVAs are (2, 87). AD = Alzheimer's disease; ANOVA = analysis of variance;

RWR = retention-weighted recall.

** $p < .001$.

Results and Discussion

Figure 1 shows that the serial position curves were similar for the two AD groups compared with controls. The results of one-way between-groups ANOVA comparing total recall, total RWR scores, delayed recall, total primacy, and total recency between controls, and the Australian AD and Korean AD groups are summarized in Table 2. Post hoc comparisons indicated the control group's scores were significantly better than scores from the Australian and Korean AD groups on measures of total recall, delayed recall, total RWR score, primacy, and recency. Figure 2 shows the magnitude of impairment of Australian and Korean AD groups on each outcome measure of the ISLT relative to matched healthy Australian controls. Figure 2 also shows that in the AD groups, the magnitude of impairment in recency was less than in primacy.

A repeated-measures ANOVA was then conducted for total recall and RWR. For total recall, repeated-measures ANOVA indicated a significant interaction between group (controls, Australian AD group, and Korean AD group) and trial (three

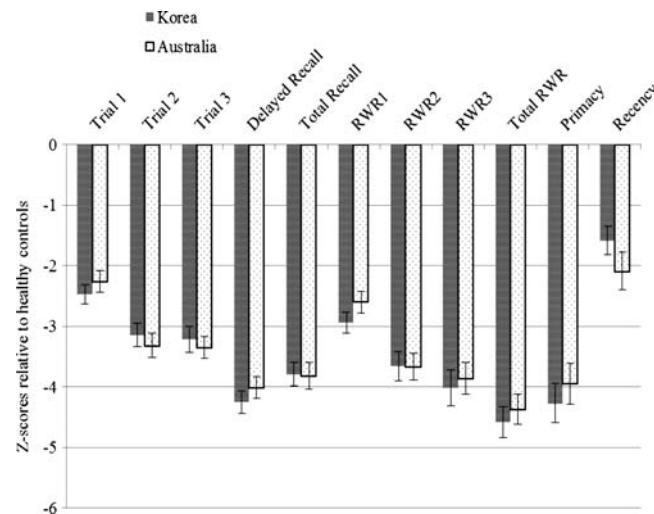


Fig. 2. Performance of Australian and Korean AD patients on each outcome measure of the ISLT relative to healthy Australian controls, Study 1. Error bars indicate 95% confidence intervals.

learning trials), Wilks' $\Lambda = 0.80$, $F(4, 172) = 5.11$, $p = .001$, Bonferroni adjusted *post hoc* pairwise comparisons indicated that total recall in the controls was better than that in the Australian and Korean AD groups (both p 's $< .001$). Total recall scores in the Australian and Korean AD groups did not differ, $p = 1.00$. The interaction reflected that the magnitude of the difference in total recall scores between the controls and the two AD groups increased across trials. This is illustrated by the increasing effect sizes for recall scores across Trials 1, 2, and 3 and delayed recall trial in Fig. 2.

For RWR, repeated-measures ANOVA indicated a significant interaction between group and trial, Wilks' $\Lambda = 0.80$, $F(4, 172) = 5.17$, $p = .001$, and significant main effects for group, $F(2, 87) = 136.16$, $p < .001$, and trials, Wilks' $\Lambda = 0.37$, $F(2, 86) = 73.09$, $p < .001$. The Bonferroni adjusted *post hoc* pairwise comparisons indicated that total RWR scores in controls were better than that in the Australian and Korean AD groups (both p 's $< .001$). Total RWR scores in the Australian and Korean AD groups did not differ, $p = 1.00$. The interaction reflected that the magnitude of the difference in RWR between the controls and the two AD groups increased as the number of trials increased (see increasing effect sizes for RWR across Trials 1, 2, and 3 on Fig. 2). Lastly, Pearson's correlations indicated total recall scores correlated significantly with delayed recall ($r = .67$, $p < .001$) and total RWR scores ($r = .85$, $p < .001$); and that delayed recall scores correlated significantly with total RWR scores ($r = .85$, $p < .001$).

Taken together, results of Study 1 indicate that AD-related memory impairment in Australian and Korean adults with AD, tested on their respective versions of the ISLT, was comparable on all of the outcome measures. We next evaluated the reliability of the ISLT measures and the magnitude of change in performance over very short (i.e., 1 week) and long (i.e., 1 year) intervals in Study 2.

Study 2

Method

Participants. Three groups of patients (30 Canadian-English speakers, 30 Korean speakers [from Study 1] and 10 French speakers) with mild AD were recruited and assessed on alternate forms of the ISLT on two occasions 7 days apart. All participants were recruited from major metropolitan health centers in Vancouver (Canada), Seoul (Korea), and Paris (France). The exclusion/inclusion criteria for Study 2 were the same as those applied in Study 1. None of the Canadian-English- or French-speaking participants had been tested with the ISLT previously. Demographic and clinical characteristics of each group is shown in Table 1. The AD groups did not differ significantly on age, education, MMSE score, or depression. From this group, 20 Korean-, 20 Canadian-English-, and 10 French-speaking patients with AD agreed to reassessment 12 months after the second baseline. The MMSE score in the 50 AD participants who completed both assessments declined from 22.3 ($SD = 5.8$) at baseline to 19.8 ($SD = 6.3$) at 12 months, $t(49) = 3.15$, $p < .001$, $d = 0.38$. All of the Canadian and Korean patients with AD had been born in their country of residence. All of the Canadian speakers were Caucasian and all of the Korean speakers were Asian. Of the French sample, all were native French speakers and 90% had been born

in France. One French participant had been born in Africa but had lived in France for 58 years. All data collected from participants complied with the regulations of institutional research and ethics committees.

Measures. The procedure outlined in Lim and colleagues (2009) was used to adapt items on the ISLT to ensure cultural and linguistic validity in Canadian-, Korean-, and French-speaking populations. These words were added to the ISLT software, and the standard instructions were translated to the appropriate language. Native speakers of each respective language administered the appropriate wordlist to respective participants. Participants were tested twice at baseline (with a week-long interval between each baseline test) and once at a 12-month follow-up.

Data analysis. Average measure intraclass correlation coefficients (ICCs) were computed for each ISLT performance measure between the two baseline assessments in the total group and for the Canadian and Korean groups. Test–retest reliability was not computed for the French group because it was too small to provide sufficient power for statistical analysis. The magnitude and statistical significance of any change in performance from the first to the second baseline assessments was also computed using a series of paired-samples *t*-tests. Means and standard deviations of the change scores are shown under “practice effects” in Table 3. The standard error of measurement for these change scores are also shown in Table 3. To determine whether annual change in ISLT performance was equivalent between language groups, performance measures with the highest test–retest reliability were used in order to minimize the Type 1 error rate and to optimize sensitivity to change. Thus, the total number of words recalled, total RWR score, and delayed recall score at the second baseline and 12-month assessments were submitted to separate between-subjects analyses of covariance (ANCOVAs), with participants’ country of origin entered as a fixed factor, and participants’ baseline score entered as a covariate. The magnitude of change over time was then determined by first combining the data from the three different language groups and subjecting the measures of total recall, total RWR, and delayed recall to a series of paired samples *t*-tests. One-year change scores were then computed for the different language groups.

Results and Discussion

ICCs indicated that the test–retest reliability of the ISLT performance measures in Canadian-English and Korean patients with mild AD were generally similar across the two different cultural contexts, and when data from French, Canadian, and Korean participants were combined, measures of total recall, delayed recall, and total RWR scores provided the highest

Table 3. Summary of average measure ICCs on each ISLT outcome measure in Canadian, French, and Korean AD groups at baseline; change in mean (*SD*) number of words recalled and standard error (*SE*) of this change in the second baseline assessment in relation to the first baseline assessment in Canadian, Korean, and French AD groups (practice effects)

Measure	Canadian ICC AD (<i>N</i> = 30)	Korean ICC AD (<i>N</i> = 30)	Canadian-Korean-French ICC AD (<i>N</i> = 70)	Canadian-Korean-French practice effects (<i>N</i> = 70)	
				<i>M</i> (<i>SD</i>)	SE change
Recall					
Trial 1	0.46	0.39	0.48*	0.01 (1.56)	0.19
Trial 2	0.70**	0.59*	0.63**	0.37 (1.33)*	0.16
Trial 3	0.58*	0.74**	0.66**	0.46 (1.49)*	0.18
Total	0.76**	0.72**	0.78**	0.86 (3.05)*	0.36
Delayed	0.65**	0.89**	0.66**	0.27 (1.18)	0.14
RWR score					
Trial 1	0.55*	0.42	0.49*	−0.16 (12.74)	1.52
Trial 2	0.72**	0.56*	0.60**	3.19 (11.58)*	1.38
Trial 3	0.42	0.68**	0.54**	2.49 (14.02)	1.68
Total	0.76**	0.65**	0.67**	5.51 (27.29)	3.26
Primacy					
Total	0.66**	0.40	0.54**	0.33 (2.73)	0.33
Recency					
Total	0.69**	0.10	0.44*	0.37 (2.64)	0.32

Notes: ICC = intraclass correlation coefficient; AD = Alzheimer’s disease;
RWR = retention-weighted recall; SE = standard error of the change score.
**p* < .05.
***p* < .001.

test–retest reliability (Table 3). These three outcome measures were therefore used to compare rates of change in memory between the different cultural groups.

Table 4 summarizes the means of the total number of words recalled, total RWR score, and delayed recall score at each assessment period. For total recall scores, ANCOVAs revealed that with baseline variability controlled statistically, there was no significant difference in performance at the 12-month assessment between the three language groups, $F(2, 46) = 0.068$, $p = .93$, partial $n^2 = 0.003$. For delayed recall scores, a second ANCOVA indicated that with the baseline variability controlled statistically, there was no significant difference in performance at the 12-month assessment between the three language groups, $F(2, 46) = 0.66$, $p = .52$, partial $n^2 = 0.028$. For RWR scores, a third ANCOVA indicated that with the baseline variability controlled statistically, there was no significant difference in performance at the 12-month assessment between the three language groups, $F(2, 46) = 0.13$, $p = .88$, partial $n^2 = 0.006$.

Given that there was no difference in performance between different language AD groups at the 12-month assessment, the magnitude of change from baseline to 12 months was evaluated by comparing performance between these two assessments in the entire AD group ($n = 50$). Paired samples t -tests indicated statistically significant declines from the baseline to the 12-month assessment for the total recall scores, $t(49) = 4.04$, $p < .001$, $d = 0.62$, delayed recall score, $t(49) = 3.65$, $p < .001$, $d = 0.52$, and total RWR, $t(49) = 4.74$, $p < .001$, $d = 0.67$. Despite the absence of an effect of language on the rate of change, group means and standard deviations were computed for the individual language groups, as well as for the entire AD group to illustrate the nature of change in performance (Table 4).

General Discussion

Results from these two studies suggest that the ISLT can reliably detect AD-related verbal learning and memory impairment in people from different cultures. In Study 1, English-speaking adults from Australia with mild AD and Korean-speaking adults from Korea with mild AD demonstrated equivalent levels of impaired performance on the ISLT. Consequently, data from the Korean AD group could be compared meaningfully with that for the Australian controls (Fig. 2). The comparison with a standard reference group demonstrated that English-speaking and Korean-speaking adults with mild AD were impaired on all ISLT performance measures. The nature of this impairment was equivalent to that observed previously in AD for other verbal list learning tests with patients recalling fewer words than controls on each trial; forgetting more words than controls after a brief delay; showing lower primacy than recency effects; and consequently obtaining lower RWR scores than controls (e.g., HVLT-R, Brandt & Benedict, 2001; Buschke et al., 2006; CVLT, Delis et al., 1987; Jungwirth et al., 2009; RAVLT, Rey, 1964; Thompson et al., 2011). The magnitudes of impairment in verbal learning and memory performance observed was, by convention, very large (e.g., control standardized scores ≥ 3 , Fig. 2). Furthermore, when words recalled were expressed as a serial position curve (Fig. 1), both Korean-speaking and English-speaking AD groups showed similar performance with fewer words recalled than healthy controls at the start (i.e., primacy) and middle of the list relative to the end of the list (i.e., recency). Direct comparison of primacy and recency effects showed that both were lower in the AD groups than in controls. This serial position effect has been observed in previous studies of AD patients and has been inferred to occur as a result of the challenge for AD patients to retain newly learned information (e.g., Bayley et al., 2000; Buschke et al., 2006).

Study 2 evaluated the ability of the ISLT to detect change in AD-related memory impairment in different cultural groups over both a short (i.e., 1 week) and long (i.e., 12 months) interval. Analysis of performance over 1 week showed the test–retest reliability of the summary ISLT outcome measures (i.e., total trials, total RWR, delayed recall) to have good reliability (e.g., $r > .70$). These estimates of reliability are consistent with those reported for other verbal list learning tests in AD (e.g., HVLT-R, Brandt & Benedict, 2001; CVLT, Delis et al., 1987; CVLT-II, Delis, Kramer, Kaplan, & Ober, 2000). More importantly, estimates of reliability were comparable for the Canadian-English and Korean AD groups. Improvement in performance

Table 4. Group mean performance on main ISLT outcome measures for Canadian-English, Korean, and French AD groups on baseline and 12-month follow-up, Study 2

	Canadian-English AD ($N = 20$)		Korean AD ($N = 20$)		French AD ($N = 10$)		Total AD ($N = 50$)	
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Total Recall	10.85 (4.78)	8.50 (4.73)	10.50 (4.36)	8.60 (5.25)	10.20 (3.19)	8.23 (4.49)	11.01 (3.52)	8.86 (4.48)
Delayed Recall	1.10 (1.48)	0.40 (0.99)	1.70 (1.95)	1.00 (1.56)	1.10 (1.20)	0.45 (0.50)	1.17 (1.52)	0.74 (1.33)
RWR score	65.50 (33.48)	44.15 (31.83)	60.45 (33.87)	44.95 (36.79)	58.70 (20.97)	40.74 (23.13)	62.36 (27.90)	43.96 (29.64)

Notes: AD = Alzheimer's disease; RWR = retention-weighted recall.

over the short term (i.e., a practice effect) was observed only for the recall measures on Trials 2 and 3 and for the total recall measure and Trial 2 RWR. That said, in each instance, the magnitude of the practice effect observed was very small (i.e., <1 word) and for the short retest interval, this estimate is probably close to maximal.

Performance measures from the ISLT that showed the highest test–retest reliability (i.e., total recall, total RWR, and delayed recall scores) were used to compare the rate of change in verbal learning and memory performance over 1 year in the different AD cultural groups. Although these summary measures were correlated, none was close to 1, which suggests that these scores do reflect different aspects of verbal learning and memory. Each of these measures also possesses a good theoretical basis for use in assessing AD-related verbal learning and memory performance (Huntley & Howard, 2009). Results of Study 2 also showed that performance over time on the ISLT measures was not affected by the language of the AD groups (Canadian-English, French, and Korean). Furthermore, consistent with previous studies that examined change over time on similar tasks of verbal list learning (e.g., HVLT, CVLT; Salmon & Bondi, 2009; Stavitsky et al., 2006), a significant decline in a moderate magnitude (e.g., ~0.5 SD) was observed to have occurred over a 12-month interval for the total recall, total RWR, and delayed recall scores.

Taken together, results of the two studies reported here indicate that the ISLT is reliable and can detect AD-related memory impairment comparably in people from different cultural groups. We believe that this equivalence in sensitivity to AD-related verbal learning and memory impairment in different cultures is related to the design of the ISLT, which overcomes many of the limitations associated with translating stimuli from verbal list learning tests into languages and cultures other than English. In particular, concerns regarding the equivalence and relevance of such instruments between cultures are raised, when the direct translation of stimulus words is attempted (e.g., Sziklas & Jones-Gotman, 2008) or when word matching strategies are used to establish equivalence between two linguistically different verbal memory tests (e.g., Agranovich & Puente, 2007). Results of the current study, as well as past studies (Lim et al., 2009; Thompson et al., 2011), demonstrate that the use of the shopping list metaphor and the selection of words describing foodstuffs that are common in the intended cultural group helps to ensure that the stimuli used are known in the target culture and possess equivalent imaginability and word frequency.

Consistent with past research, ISLT performance measures of total recall, delayed recall, and total RWR demonstrated the largest magnitude to memory impairment (standard scores >3; Buschke et al., 2006; Gaines, Shapiro, Alt, & Benedict, 2006; Jungwirth et al., 2009; Thompson et al., 2011). The RWR score was suggested by Buschke and colleagues (2006) as a way to improve the sensitivity of the verbal memory test to AD-related memory impairment by combining the indices of total recall and primacy, both of which are impaired in AD. The RWR was computed for a single trial 10-word learning task and was observed to have sensitivity to AD-related memory impairment greater than that observed for the total recall score. In our previous study, we also found that the single trial RWR was more sensitive to AD-related memory impairment than free recall but only for the first trial of the ISLT (Thompson et al., 2011). However, in the current study, the total RWR score produced estimates of test–retest reliability, as well as effect sizes of differences in memory performance between mild AD and controls, and magnitude of deterioration in performance over 12 months that were equivalent to the total recall score. Despite the improved characteristics of the total RWR relative to a single trial RWR, for no comparisons was the total RWR score superior to the total recall score. This equivalence of the RWR and total recall is also consistent with the results of a recent study which reported that the sensitivity of the total RWR score was not better than that of the total recall score in discriminating between adults with AD and healthy controls on the verbal list learning test from the Neuropsychological Assessment Battery (Howieson et al., 2011). Interestingly, this equivalence of sensitivity extended to include the correspondence analysis measure developed by Shankle and colleagues (2005), where a statistical program maximizes the correlation between groups and recall by taking into account both words recalled and not recalled, and the short-term memory penalty. These data suggest that the measures used conventionally to characterize performance on verbal list learning tests have optimal discriminative ability and that the combinations of scores or statistical manipulation of outcome measures suggested to date do not further improve the sensitivity of verbal list learning to assessing AD-related memory impairment.

In addition to developing stimulus sets that are appropriate for the language of interest in cross-cultural studies, it is also important to consider the extent to which the people being assessed are familiar with testing and assessment more generally (Ardila, 2005; Reynolds, 2000). We have argued that the shopping list format of the ISLT is more suitable for groups where formal memory assessment is not common, because the format is consistent with a common activity of daily living (Lim et al., 2009). Given that all of the participants in this study were recruited from clinical sites and had undergone formal neuropsychological and clinical assessment irrespective of the culture in which they lived, the value of the shopping list format for making valid cross-cultural assessments of verbal memory in individuals who have limited test taking experience remains to be evaluated. One way to address this may be to administer the ISLT to individuals in an environmental setting where Western European conventions of test taking are not common, such as a rural area or a non-Western country, and then compare their performance to individuals familiar with the test taking procedures. In this circumstance, it is expected that individuals who have low test-taking experience would not perform poorly on the ISLT, as the shopping list format should be familiar to

most individuals. Results of Study 2 show that performance on the ISLT outcome measures remained stable, despite a retest conducted at a relatively short retest interval (1 week). While practice effects were observed, they were generally very small (<1 word). Thus, in cultures where people may have difficulty performing the ISLT, it may be possible to familiarize examinees with the ISLT through repeated assessments. It would be useful, however, to first investigate the stability of the ISLT performance measures over shorter (e.g., hours) retest intervals.

There were several limitations associated with the current study. First, the sample studied here consisted only of a small group of adults with AD. Thus, these findings require replication in larger groups of persons with AD, as well as other disease groups with known reductions in verbal learning and memory. Second, in the current study, we chose to use an Australian normative group as a single reference sample. However, before the ISLT can be used clinically or in research studies investigating populations who speak languages other than Australian English, culturally appropriate normative ranges are necessary. Furthermore, equivalence in performance between healthy older adults from the different cultural and linguistic groups needs to be shown before we can conclude that the ISLT has equivalent sensitivity to AD-related memory impairment in different cultural groups. Third, the use of the shopping list format also means that the resultant stimulus sets belong to a single semantic category. Most other verbal list learning tests deliberately include words that belong to three or four semantic categories (e.g., HVLT-R, CVLT-II) or choose words that are semantically unrelated (e.g., RAVLT). Nevertheless, results of the current study and from previous studies suggest that individuals with mild AD show very large impairments on the ISLT despite all of the stimuli being derived from a single semantic category. For example, with the advantage of a very large control group, [Thompson and colleagues \(2011\)](#) showed that many of the ISLT performance measures possessed sensitivity of 95% (at 90% specificity) to detecting verbal memory impairment in mild AD. Furthermore, the nature and magnitude of differences between the performance of controls and people with mild AD observed for the ISLT were consistent with those observed for verbal list learning tests that employ either three or four semantic categories ([Salmon & Bondi, 2009](#); [Stavitsky et al., 2006](#)). While the utility of the use of semantic categories to understand verbal memory impairment in mild AD is not the subject of this study, our results and the observation that no verbal list learning test is clearly more sensitive to AD-related impairment than the other suggest that semantic categories may not be important to the assessment of verbal episodic memory in mild AD. This is being an area for further examination.

It is well established that culturally appropriate and valid tests are important in the assessment of cognitive functions ([Ardila, 2005](#); [Greenfield, 1997](#); [Reynolds, 2000](#)). Tests developed in North America or Western Europe cannot be easily translated for use in other cultures, specifically due to issues surrounding relevance of stimuli used and test-taking experience of individuals in other cultures. Differences in test performance that result from the use of such tests in comparing performance of individuals from other cultures may be mediated by several factors and may pose limitations in conclusions drawn for use in diagnosis and intervention. The ISLT provides a paradigm that attempts to address these limitations and may serve as a useful tool in the neuropsychological assessment of verbal learning and memory performance in specific clinical populations.

Conflict of Interest

P.M. is a full-time employee of CogState Ltd, the company that provides the International Shopping List Test. R.H.P., P.J.S., and D.D. are consultants to this company. Y.Y.L. does not declare any conflicts of interest.

Appendix

Table A1 shows the eight lists of the Australian version of the ISLT used in the current study and in [Thompson and colleagues \(2011\)](#). Table A2 gives the group mean performance on each outcome measure from the ISLT in the subgroups of the [Thompson and colleagues \(2011\)](#) sample who completed the different ISLT lists detailed in Table A1.

Table A1. Eight versions of the Australian ISLT used in the Australian sample in the current study and in Thompson and colleagues (2011)

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
Tea	Coffee	Beer	Water	Juice	Wine	Cordial	Lemonade
Potatoes	Tomatoes	Onions	Carrots	Peas	Beans	Pumpkin	Cauliflower
Apples	Oranges	Bananas	Lemons	Grapes	Pears	Apricots	Strawberries
Chocolate	Lollies	Liquorice	Popcorn	Chips	Mints	Nuts	Sultanas
Pastie	Pizza	Hamburger	Chips	Soup	Dim sim	Noodles	Pie
Eggs	Cheese	Yoghurt	Custard	Milk	Ice cream	Butter	Cream
Cake	Bread	Donuts	Croissants	Crumpets	Rolls	Pavlova	Biscuits
Mayonnaise	Vegemite	Jam	Margarine	Honey	Marmalade	Gravy	Chutney
Spaghetti	Lasagne	Porridge	Rice	Ravioli	Cereal	Macaroni	Flour
Pickles	Ketchup	Mustard	Parmesan	Parsley	Sauce	Syrup	Vinegar
Sugar	Salt	Pepper	Vanilla	Chilli	Garlic	Oregano	Ginger
Steak	Turkey	Sausage	Cabana	Chicken	Salami	Bacon	Ham

Table A2. Group mean (*SD*) performance on each list of the Australian ISLT computed here from the normative group described in Thompson and colleagues (2011)

Measure	List 1 (<i>n</i> = 19)		List 2 (<i>n</i> = 18)		List 3 (<i>n</i> = 19)		List 4 (<i>n</i> = 21)		List 5 (<i>n</i> = 21)		List 6 (<i>n</i> = 19)		List 7 (<i>n</i> = 20)		List 8 (<i>n</i> = 19)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Trial 1	6.8	1.6	6.5	1.8	6.9	1.7	6.2	2.1	6.4	1.9	6.0	1.6	6.9	1.3	6.2	1.8
Trial 2	9.4	1.1	9.1	1.3	9.2	1.3	9.5	1.7	9.4	1.6	9.2	1.5	9.8	1.3	9.5	1.7
Trial 3	10.4	1.6	10.7	1.4	10.7	1.3	10.7	1.6	10.3	1.1	10.7	1.6	10.1	1.2	10.9	1.3
Total Recall	26.6	3.9	26.3	3.2	9.5	2.6	26.4	3.7	26.1	4.1	25.9	3.8	26.8	3.9	26.6	3.6
Delayed Recall	9.1	2.3	9.2	2.1	26.6	4.1	9.0	2.2	9.7	2.4	9.5	2.0	9.6	2.5	10.9	1.3
RWR 1	49.7	12.8	47.5	12.1	50.4	11.9	45.3	12.3	46.8	11.9	43.9	13.1	50.4	12.9	45.3	12.4
RWR 2	70.2	15.1	68.0	14.1	67.2	16.2	70.9	14.2	70.2	14.7	68.7	15.9	73.2	15.4	70.9	15.1
RWR 3	79.2	17.1	81.5	17.9	81.5	16.7	81.5	17.8	78.4	18.3	81.5	16.8	76.9	17.1	83.0	18.4
Total RWR	199.1	21.6	197.0	23.6	199.1	25.8	197.7	23.1	195.4	27.1	194.0	26.8	200.5	29.5	199.3	28.1
Primacy	8.3	1.1	9.6	1.5	9.1	1.8	9.0	1.1	8.6	1.9	9.8	1.6	8.1	1.8	8.2	1.3
Recency	8.1	1.8	8.2	1.5	9.2	1.6	8.1	1.7	9.1	2.1	9.3	1.1	8.3	1.6	8.3	1.1

References

Agranovich, V., & Puente, A. E. (2007). Do Russian and American normal adults perform similarly on neuropsychological tests? Preliminary findings on the relationship between culture and test performance. *Archives of Clinical Neuropsychology*, 22, 273–282.

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

Ardila, A. (2007). Toward the development of a cross-linguistic naming test. *Archives of Clinical Neuropsychology*, 22 (3), 297–307.

Baddeley, A. (1992). Working memory. *Science*, 255 (5044), 556–559.

Bayley, P. J., Salmon, D. P., Bondi, M. W., Bui, B. K., Olichney, J., Delis, D. C., et al. (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 Neuropsychology Society*, 6, 290–298.

Bock, M., & Klinger, E. (1986). Interaction of emotion and cognition in word recall. *Psychological Research*, 48, 99–106.

Brandt, J., & Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test-Revised*. Odessa, FL: PAR.

Buschke, H., Sliwinski, M. J., Kuslanski, G., Katz, M., Verghese, J., & Lipton, R. B. (2006). Retention weighted recall improves discrimination of Alzheimer’s disease. *Journal of the International Neuropsychological Society*, 12, 436–440.

Chow, T. W., Liu, C. K., Fuh, J. L., Leung, V. P. Y., Tai, C. T., Chen, L. W., et al. (2002). Neuropsychiatric symptoms of Alzheimer’s disease differ in Chinese and American patients. *International Journal of Geriatric Psychiatry*, 17 (1), 22–28.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult version*. San Antonio: The Psychological Corporation, Harcourt Brace & Company.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test—Second Edition (CVLT-II)*. San Antonio, TX: The Psychological Corporation.

Demers, P., Robillard, A., Lafliche, G., Nash, F., Heyman, A., & Fillenbaum, G. (1994). Translation of clinical and neuropsychological instruments into French: The CERAD experience. *Age and Aging*, 23 (6), 449–451.

- Dick, M. B., Dick-Muehlke, C., & Teng, E. L. (2006). Assessment of cognitive status in Asians. In G. Yeo, & D. Gallagher-Thompson (Eds.), *Ethnicity and the dementias* (2nd ed., pp. 55–69). Washington DC: Taylor and Francis Group, LLC.
- Frederickson, J., Maruff, P., Woodward, M., Moore, L., Frederickson, A., Sach, J., et al. (2010). Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*, 34, 64–75.
- Gaines, J. J., Shapiro, A., Alt, M., & Benedict, R. H. B. (2006). Semantic clustering indexes for the Hopkins Verbal Learning Test-Revised: Initial exploration in elder control and dementia groups. *Applied Neuropsychology*, 13, 213–222.
- Glanzer, M. (1972). Storage mechanisms in recall. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 5, pp. 129–193). New York: Academic Press.
- Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist*, 52, 1115–1124.
- Greenway, M. C., Lacritz, L. H., Binegar, D., Weiner, M. F., Lipton, A., & Cullum, C. M. (2006). Patterns of verbal memory performance in mild cognitive impairment, Alzheimer's disease and normal aging. *Cognitive and Behavioural Neurology*, 19, 79–84.
- Howieson, D. B., Mattek, N., Seyle, A. M., Dodge, H. H., Wasserman, D., Zitzelberger, T., et al. (2011). Serial position effects in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, 33 (3), 292–299.
- Huntley, J. D., & Howard, R. J. (2009). Working memory in early Alzheimer's disease: A neuropsychological review. *International Journal of Geriatric Psychiatry*, 25 (2), 121–132.
- Jungwirth, S., Zehetmayer, S., Bauer, P., Weissgram, S., Tragl, K. H., & Fischer, P. (2009). Prediction of Alzheimer dementia with short neuropsychological instruments. *Journal of Neural Transmission*, 116, 1513–1521.
- Kalaria, R. N., Maestre, G. E., Arizaga, R., Friedland, R. P., Galasko, D., Hall, K., et al. (2008). Alzheimer's disease and vascular dementia in developing countries: Prevalence, management and risk factors. *The Lancet Neurology*, 7 (9), 812–826.
- Lee, J. H., Lee, K. U., Lee, D. Y., Kim, K. W., Jhoo, J. H., Kim, J. H., et al. (2002). Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease assessment packet (CERAD-K): Clinical and neuropsychological assessment batteries. *Journal of Gerontology: Psychological Sciences*, 57 (1), 47–53.
- Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory assessment. *Behavior Research Methods*, 41 (4), 1190–1200.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, 34, 939–944.
- Mitrushina, M., Boone, K. B., Razani, J., & D'elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). New York: Oxford University Press.
- Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice*. Mahwah, NJ: Erlbaum.
- Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behaviour*, 7, 41–49.
- Perry, R. J., Watson, P., & Hodges, J. R. (2000). The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: Relationship to episodic and semantic memory impairment. *Neuropsychologia*, 38 (3), 252–271.
- Pietrzak, R. H., Olver, J., Norman, T., Piskulic, D., Maruff, P., & Snyder, P. J. (2009). A comparison of the CogState Schizophrenia Battery and the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Battery in assessing cognitive impairment in chronic schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, 31, 848–859.
- Rey, A. (1964). *L'examen Clinique en Psychologie*. Paris: Press Universitaire de France.
- Reynolds, C. R. (2000). Methods for detecting and evaluating cultural bias in neuropsychological tests. In E. Fletcher-Janzen, T.L. Strickland, & C.R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 249–285). New York: Kluwer Academic.
- Salmon, D. P., & Bondi, M. W. (2009). Neuropsychological assessment of dementia. *Annual Review of Psychology*, 60, 257–282.
- Shankle, W. R., Romney, A. K., Hara, J., Fortier, D., Dick, M. B., Chen, J. M., et al. (2005). Methods to improve the detection of mild cognitive impairment. *Proceedings of the National Academy of Sciences of the USA*, 102 (13), 4919–4924.
- Shirayev, E., & Levy, D. (2007). *Cross-cultural psychology: Critical thinking and contemporary applications*. New York: Pearson Education.
- Stavitsky, K., Brickman, A. M., Scarmeas, N., Torgan, R. L., Tang, M. X., Albert, M., et al. (2006). The progression of cognitive, psychiatric symptoms, and functional abilities in dementia with Lewy bodies and Alzheimer disease. *Archives of Neurology*, 63, 1450–1456.
- Sziklas, V., & Jones-Gotman, M. (2008). RAVLT and nonverbal analog: French forms and clinical findings. *Canadian Journal of Neurological Sciences*, 35, 323–330.
- Thompson, T. A. C., Wilson, P., Snyder, P. J., Pietrzak, R. H., Darby, D., Maruff, P., et al. (2011). Sensitivity and test-retest reliability of the International Shopping List Test in assessing verbal learning and memory in mild Alzheimer's disease. *Archives of Clinical Neuropsychology*, 26 (5), 412–424.
- Wilson, B., Cockburn, J., Baddeley, A., & Hiorns, R. (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. *Journal of Clinical and Experimental Neuropsychology*, 11, 855–870.
- Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1991). First-last names and the grocery list selective reminding test: Two computerized measures of everyday verbal learning. *Archives of Clinical Neuropsychology*, 6, 287–300.