
REHAB SERIES

Cognitive rehabilitation in the elderly: Effects on memory

FERGUS I.M. CRAIK,^{1,2} GORDON WINOCUR,^{1–4} HEATHER PALMER,¹ MALCOLM A. BINNS,¹
MELISSA EDWARDS,¹ KRISTEN BRIDGES,¹ PETER GLAZER,¹ RAYONNE CHAVANNES,¹
AND DONALD T. STUSS^{1,2,5,6}

¹Rotman Research Institute, Baycrest, Toronto, Ontario, Canada

²Department of Psychology, University of Toronto, Ontario, Canada

³Department of Medicine (Psychiatry), University of Toronto, Ontario, Canada

⁴Department of Psychology, Trent University, Peterborough, Canada

⁵Department of Medicine (Neurology), University of Toronto, Ontario, Canada

⁶Department of Medicine (Rehabilitation Sciences), University of Toronto, Ontario, Canada

(RECEIVED November 11, 2005; FINAL REVISION August 18, 2006; ACCEPTED August 18, 2006)

Abstract

This study reports the effects of a 12-week multimodal cognitive rehabilitation training program on memory performance in two groups of older adults. In the Memory Training module, participants were instructed on the nature of memory and how to improve memory performance; internal and external strategies were described and practiced over the training sessions. Memory performance was assessed by four tests: Alpha Span, Brown-Peterson, Hopkins Verbal Learning Test – Revised (HVLTR), and Logical Stories. One group received training on entry into the study (Early Training Group, ETG), the other after a 3-month delay (Late Training Group, LTG). The results showed no training-related improvement in working memory (Alpha Span), primary memory (Brown-Peterson, HVLTR), or recognition memory (HVLTR). While the most direct analyses of a training effect (analyses of covariance) rarely demonstrated significant effects, exploratory analyses provided some evidence for a training benefit in several measures of secondary memory (Logical Stories; HVLTR) and strategic processing (Brown-Peterson; Logical Stories; HVLTR). Positive results were largely restricted to the ETG, possibly because the LTG lost motivation as a consequence of their delayed training. The results need to be treated with caution, but are promising for the rehabilitation of memory functioning in older adults (*JINS*, 2007, 13, 132–142.)

Keywords: Aging, Memory, Neuropsychological tests, Treatment outcome, Cognition, Frontal lobe

INTRODUCTION

Memory loss is probably the most consistent feature of cognitive impairment associated with brain damage and normal aging. It is certainly one of the most frequent complaints of the elderly and can seriously affect a person's quality of life. Nevertheless, until recently, retraining programs have been directed mainly at the memory problems of brain-damaged populations, with few programs designed specifically for the elderly (for review, see Glisky & Glisky, 1999). For these reasons, a priority objective in designing a rehabilitation program suitable for older adults was to improve memory performance in our participants.

Our approach was guided by encouraging results from the limited research into memory training in the elderly (Glisky & Glisky, 1999). For example, Ball and colleagues (2002), in the large ACTIVE study, compared three different cognitive training interventions (memory, reasoning, and speed of processing) and included a no-contact control group. There was improvement in each targeted cognitive ability, although the percentage of those who demonstrated reliable improvement after training was less in memory (26%) than in reasoning (74%) and speed (87%). Other studies demonstrated positive effects after memory training (e.g., Scogin et al., 1985), with long-term retention of improvement lasting from months (e.g., Sheikh et al., 1986) to over 3 years (Stigsdotter Neely & Bäckman, 1993).

Memory is a multidimensional construct, with different types of memory affected to varying degrees in old age. Our approach to rehabilitation was based on a theoretical

Correspondence and reprint requests to: Fergus I.M. Craik, Ph.D., Rotman Research Institute, Baycrest, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada. E-mail: fcraik@rotman-baycrest.on.ca

framework that differentiates two major components of remembering: one more automatic and familiarity-based, the other more controlled and recollective (Jacoby, 1991). Aging adversely affects recollection, but leaves familiarity relatively intact (Jennings & Jacoby, 1993). Evidence suggests that training improves controlled recollection, but not the more automatic familiarity aspects (Jacoby et al., 1996). In the present study, we assessed recollective memory in several tests that were also designed to measure primary and secondary memory, as well as working memory.

The term “short-term memory” is used in different ways by cognitive researchers and by clinicians, so we avoid it in favor of Waugh and Norman’s (1965) distinction between primary and secondary memory. In their usage, the short-term retention of information is mediated partly by the small amount of information that is maintained continuously in mind, and partly by further information that was perceived and encoded, but then dropped out of conscious awareness, and so must be retrieved to be recalled. These components of short-term retention are referred to as primary memory and secondary memory, respectively. Primary memory is affected very little by normal aging (Craik, 1968) or even by amnesia (Baddeley & Warrington, 1970), so it seemed unlikely that the present rehabilitation procedures would benefit this form of memory. On the other hand, secondary memory is the type involved in real-life episodic memory whose performance depends on the efficiency of encoding and retrieval processes, both of which should be sensitive to rehabilitation.

Working memory (Baddeley & Hitch, 1974; Engle & Kane, 2004) is similar to primary memory in that both terms refer to a small amount of material held in conscious awareness. The difference is that, whereas primary memory involves the verbatim reproduction of presented material, working memory involves the manipulation and transformation of the material held in mind. The efficiency of working memory operations, particularly if the task requires greater involvement of the “central executive,” is crucial to a wide range of our cognitive behaviors.

In the present study, working memory was assessed by the Alpha Span Test (Craik, 1986), and primary memory was measured both by the Brown-Peterson Test (Floden et al., 2000) and the Hopkins Verbal Learning Test – Revised (HVLT-R; Benedict et al., 1998; Brandt, 1991). Measures of secondary memory were obtained from the Brown-Peterson Test, the HVLT-R, and also from the Logical Stories Test (Dixon et al., 1989). These tests include measures of immediate and delayed recall, recognition, as well as strategic processing (subjective organization and semantic clustering).

METHODS

Participants and Design

Forty-nine healthy independent-living older (71–87 years old) adults, who had subjective complaints of cognitive or

memory impairment participated (see Stuss et al., 2007, for details). The volunteers who met inclusion criteria were divided into an Early Training Group (ETG; $N = 29$, 14 men) and a Late Training Group (LTG; $N = 20$, 8 men). A blocked randomization procedure was used, constrained by the need to equalize the groups with respect to the Mini-Mental State Examination (MMSE; Folstein et al., 1975), education, sex, and age. The groups were well matched on demographic and neuropsychological variables, with one exception [Logical Memory immediate recall (Wechsler Memory Scale-Revised; Wechsler, 1987)], in which the ETG was better than the LTG (Stuss et al., 2007). For the ETG, training began immediately after admission into the program. The LTG served as a control, beginning training 3 months later. The study was approved by the Baycrest Research Ethics Board and conducted in accordance with the guidelines of the Helsinki Declaration.

In overview, the complete program consisted of three 4-week modules: Memory Skills Training, Goal Management Training, and Psychosocial Training. There were four testing sessions, spaced 3 months apart, referred to as Assessments A, B, C, and D: the first session (Assessment A) provided pretraining baseline measures for both groups; Assessment B followed training for the ETG but not for the LTG; the ETG received no further training before Assessment C, but the LTG received training between Assessments B and C; finally, Assessment D was a long-term follow-up, occurring for each group 6 months after training. The training sessions were administered to groups of 5–6 participants. Missed attendance was negligible, with make-up sessions provided to ensure that all participants received the same amount of training. Homework completion was monitored to maximize comparability between groups. In this study, we provide the procedures and results related to the Memory Training module.

Memory Training

The focus of the Memory Training module was on learning a variety of strategies and techniques to improve organizational and memory skills. Consistent with other modules, the first memory training session began with a review of the rehabilitation program goal, including its aims and objectives. This session was followed by an interactive lecture that highlighted the complexity of memory and its relationship to brain function. Participants learned of the different types of memory and that forgetting can occur at different rates and for different reasons. Factors that affect successful encoding and retrieval (e.g., attention, fatigue, physical limitations, pain, medication, stress) were identified. Self-awareness of individual memory slips was promoted by encouraging participants to maintain a “slips” log throughout the module. At the completion of the first session, participants were briefly introduced to the importance of external strategies and asked to log the external strategies they used throughout the following week.

The second training session continued the focus on external strategies. It began with a review and discussion of memory slips that participants logged throughout the week. The use of strategies such as date books/diaries, post-it notes, timers, and wall calendars was encouraged. Additional techniques and strategies for enhancing remembering were discussed. These included external self-talk (the process of working through a problem or situation by thinking out loud), routines and habits (placing car keys in the same location in the house ensures being able to find them when needed), and organization and planning in ways that would reduce the number of memory slips. At the conclusion of this session, participants were encouraged to use external strategies and reduce reliance on internal strategies and “just remembering.” As homework, they were asked to continue logging slips and identify external strategies that may prevent slips from occurring. In addition, participants were given a list of internal strategies, definitions, and examples as preparation for the final two memory seminars.

The primary focus of training sessions 3 and 4 was to practice a variety of internal strategies which help to encode information in a deep and meaningful way. Participants were shown a series of graphs demonstrating that deep levels of processing at encoding can significantly increase recall (Craik & Lockhart, 1972; Craik & Tulving, 1975). Building on that discussion, participants were introduced to the following six strategies: categorization, story making, visual imagery, association, motor movement, and spaced retrieval. Each strategy was discussed with examples. Participants were then provided with opportunities to practice using each strategy. As homework between sessions 3 and 4, participants were given article reading and name remembering assignments for practicing the various strategies.

Although participants were encouraged to learn and practice all strategies, it was recognized that different strategies work for different people in different situations and that combinations of strategies are commonly used. With this understanding, participants were presented with additional test opportunities in which they were instructed to use the strategy or *combination* of strategies that best suited them.

Dependent Measures

Alpha Span Test (Craik, 1986)

Four equivalent versions of the Alpha Span Test were used at Assessments A, B, C, and D. Each version of the test consisted of 14 lists of common one-syllable words. The lists varied in length from two to eight words, and there were two lists at each length. Presentation started with list-lengths of two words and proceeded with increasingly longer lists until the participant failed on both lists of the same length. On each trial, the participant’s task was to rearrange the words mentally and recall them orally in alphabetical order. The two measures taken are Total Score, the number of errorless trials (ranging from 0 to 14), and Partial Score, where participants were given 1 point for each item that

was recalled as a member of a correctly recalled adjacent pair.

Brown-Peterson Test (Floden et al., 2000, modification)

In this test, a list of four unrelated words was presented orally for the participant to recall orally after a delay of 0, 3, 18, or 36 s. The 48 words used in the 12 scored trials were high-frequency, concrete nouns of high imagery value; they were grouped arbitrarily into sets of four with no obvious associative links among members of each set. The delay intervals were filled with a rehearsal-preventing counting task, consisting of subtracting either 3, 2, or 1 from each of a series of two-digit numbers, presented visually. To equate the difficulty of the distracter task between participants, the task was first titrated to achieve a 60% accuracy rate at a 1-s presentation rate. During the titration phase, participants were presented with two-digit numbers at rates of 1.6, 1.2, and then 1.0 s per number. They were asked to complete the subtraction task to determine a condition enabling each participant to be 60% accurate at the 1-s rate.

In the test proper, short lists of four words were presented orally at a 2-s rate. Participants repeated words out loud to verify that they had heard each word correctly. Following presentation, participants were either cued immediately for recall (0 delay) or performed the subtraction task until the recall cue (a visual question mark) was presented for 2 s. The screen remained blank for an additional 18 s, allowing a total of 20 s for recall. The entire experiment consisted of three practice trials at a 3-s delay, and then three experimental trials at each of the four delay intervals, presented quasirandomly. The distractor task was presented visually at a rate of one number per second, and the amount subtracted from each number was set at the value attained in the titration phase. The same materials were presented on all four testing occasions.

The following dependent measures were analyzed:

- Secondary memory*—average correct word recall at the 36-s delay, on the assumption that primary memory is no longer available after 36 s of distracting activity.
- Primary memory*—average correct word recall over the 0-, 3-, and 18-s delays. At each delay, primary memory is calculated from the formula:

$$PM = \text{Observed recall} - SM \text{ recall}/1 - SM \text{ recall}$$

in which PM = probability of primary memory recall, SM = probability of secondary memory recall (from the 36-s condition), and observed recall is the raw probability of recall at that delay interval (see Waugh & Norman, 1965, for fuller details).

- Resource allocation*. On the assumption that optimal performance involves allocating available resources to recalling the words, we calculated two measures, termed *Performance* and *Strategy*. *Performance* is the sum of

the proportion of correct words and correct number subtractions at the 36-s delay. An increase in this measure indicates improved task performance. *Strategy* is the difference in proportions between correct words and correct numbers at the 36-s delay. An increase in this measure indicates that more resources are being allocated to recall.

Hopkins Verbal Learning Test-Revised (HVLTR)

The HVLTR (Benedict et al., 1998; Brandt, 1991) measures primary and secondary memory in addition to the rate of verbal learning over three successive learning and recall trials. A 12-item word list is presented orally at a rate of one word every 2 s, and the participant recalls as many words as possible in any order immediately after the presentation. Each list consists of three semantic categories with four words from each category. A different HVLTR form was used on each testing occasion. After the participant could no longer recall any words, the list was presented again, followed by a further recall attempt. Three successive presentations and recall tests were administered. After the third learning trial, participants were instructed to remember the words because they would be asked to recall them again later in the session. After a 30-min delay in which other cognitive tests were administered, the participant again recalled the words in any order (delayed free recall). This delayed test was followed by a recognition test consisting of the 12 target words randomly mixed with 12 distracter words, with half of the distracter words being drawn from the same semantic categories as the targets (related distracters) and the other half from other categories (unrelated distracters).

The measures include correct immediate and delayed recall, and delayed recognition. For immediate recall, the total was broken down into primary and secondary memory scores. A recalled word was counted as coming from primary memory provided that no more than seven words intervened between the word's presentation and recall (Tulving & Colotla, 1970). The remaining recalled words were judged to come from secondary memory. In addition, three different measures of strategic organization in free recall were calculated: serial ordering, semantic clustering, and subjective organization. For serial ordering, 1 point was given each time the participant recalled two correct words in the same order as presented. From this total, we subtracted the number of serially recalled pairs expected by chance, and used this chance-adjusted score in the analyses. For semantic clustering, 1 point was given for each correct word following another word from the same semantic category. Subjective organization (Tulving, 1966) was defined as the number of correct word pairs recalled together (regardless of order) that had been recalled together in the previous trial. All strategic measures were adjusted based on the number of words recalled. Word repetitions and intrusions were also recorded. For the recognition test, the principal mea-

sure was number of hits minus number of false-positive errors made to distracter items.

Logical Stories Test

Four stories were selected from the original 25-story collection created by Dixon et al. (1989). The stories were designed to be semantically and structurally homologous, describing a single concrete event in the life of an older protagonist. In each story, we used the first paragraph—eight sentences containing approximately 100 words and an average of 53.8 propositions or idea units.

Each story was read aloud to participants who were then asked to recall as much information as they could, using as many of the same words as they could remember. Recall testing took place immediately (Immediate) and again approximately 30 min later (Delayed). If participants failed to recall the story during Delayed Recall, they were prompted with the gist of the story. Participants' oral recall was recorded and later transcribed for scoring. A different story was used on each test occasion, but stories were not counterbalanced over tests; that is, the same story was given to all participants on each specific test occasion.

Scoring was in terms of recollection of the specific and general points related in each story. These story items were designated as belonging to Level 1 (general gist of the story), Level 2 (details of the story), or Level 3 (highly specific details). Performance was calculated as a percentage of the possible total at each level. Measures of Levels 1, 2, and 3 plus the total number of story items recalled were calculated for both Immediate and Delayed Recall, making a total of eight measures overall.

Statistical Analysis

The general approach to data analysis is described in the Introductory paper (Stuss et al., 2007). Performance at Assessment A was examined with analysis of variance (ANOVA) to determine whether the ETG and LTG differed on any measure before rehabilitation training. Analysis of covariance (ANCOVA) was used to examine performance at each of Assessments B and C, while statistically controlling for performance at the previous session (Assessments A and B, respectively). Repeated measures analysis of variance was used to investigate long-term performance changes between Assessments A and D. Effect sizes are presented for all significant group comparisons.

One exception to this general data analysis strategy was required for the HVLTR data. Scores from a subset of 7 ETG participants at Assessment A had to be excluded from analysis because they were allowed to write their responses as opposed to reporting them orally. In order to not suffer an unacceptable loss of subjects due to missing values at Assessment A, analysis of variance was carried out for HVLTR at Assessment B without covarying Assessment A performance.

As indicated in the Introductory paper, 5 participants in the LTG did not complete Assessments C and D due to an

outbreak of severe acute respiratory syndrome (SARS), which effectively prohibited research participants from entering the hospital.

Measures from the four memory tests fall into three general categories which were anticipated to respond differentially to this rehabilitation training program: two requiring relatively little recollection—working memory, primary memory; and one category dependent more on controlled recollection—secondary memory (including strategic processing related to efficient secondary memory). Working memory was assessed using the Alpha Span Test. Both the Brown-Peterson and Hopkins Verbal Learning Tests provided measures of primary and secondary memory. Strategic processing was directly targeted by the program, and measures of this were built into Brown-Peterson, HVLT (e.g., semantic clustering and subjective organization in HVLT), and Logical Stories.

RESULTS

Figures 1–4 show mean scores and standard errors for the participants who were tested at each session. Table 1 provides the means and standard deviations for the main measures at the four assessments. There were no significant group differences for any of the memory measures at Assessment A, indicating that the ETG and LTG were equivalent in terms of baseline memory functioning.

Alpha Span Test

ANCOVA, taking previous performance levels into account as described in the Introductory paper, were performed on the Alpha Span Test scores at Assessments B and C. These analyses failed to reveal any differences between the ETG and LTG. [ANOVA conducted on the working memory measures did not reveal significant group differences at any of

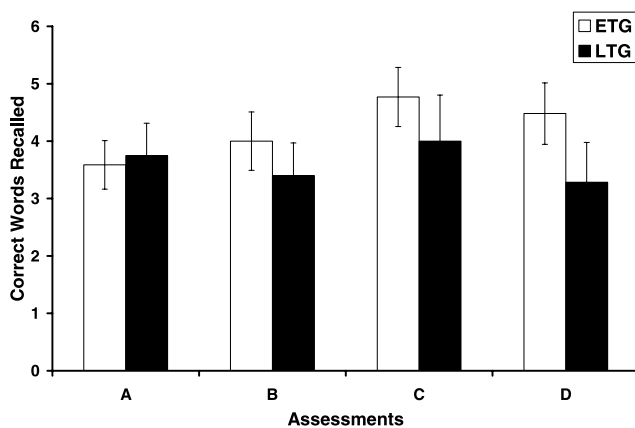


Fig. 1. Number of words recalled correctly from secondary memory (36-second delay) by the Early (ETG) and Late (LTG) Training Groups at all assessments of the Brown-Peterson Test. The maximum possible score is 12. Error bars represent the standard error of the mean.

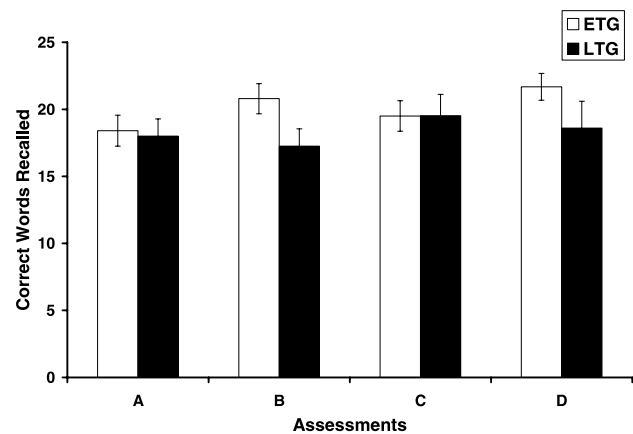


Fig. 2. Number of words recalled correctly from secondary memory out of a possible 36 (immediate recall) by the Early (ETG) and Late (LTG) Training Groups at all assessments on the Hopkins Verbal Learning Test–Revised (HVLT-R) Test. Error bars represent the standard error of the mean.

the assessments.] Moreover, a series of within-group comparisons involving participants who completed successive pairs of test sessions also failed to reveal any significant effects. It appears that memory training had no effect on this measure of working memory performance.

Brown-Peterson Test

For the primary memory measures, there were no meaningful group effects, as revealed by ANCOVA. [The only exception among the 12 primary memory measures (0-, 3-, and 18-s delays \times 4 assessments) was at 0 delay for Assessment B, where a significant difference ($p < .04$) favoring the ETG was observed. We consider this a chance finding. Additionally, ANOVA conducted on the primary memory measures did not reveal significant group differences at any of the assessments.] In the following, we report findings related to secondary memory, as assessed by the number of words recalled at the 36-s delay.

Training effects on ETG (Assessments A to B)

An ANCOVA applied to the Assessment B data did not reveal significant group effects.

Training effects on LTG (Assessments B to C)

There were no significant between-group differences at Assessment C. However, within-group comparisons across Assessments B and C showed that the LTG improved significantly on words recalled at the 36-s delay ($F_{1,14} = 10.23$, $p = .006$), while ETG did not improve to the same degree ($F_{1,25} = 3.19$, $p = .09$; Figure 1). There was also a significant improvement in the LTG's strategy score at Assessment C, relative to Assessment B ($F_{1,14} = 5.80$, $p = .03$); the ETG did not improve to the same degree ($F_{1,28} = 2.64$, $p = .12$).

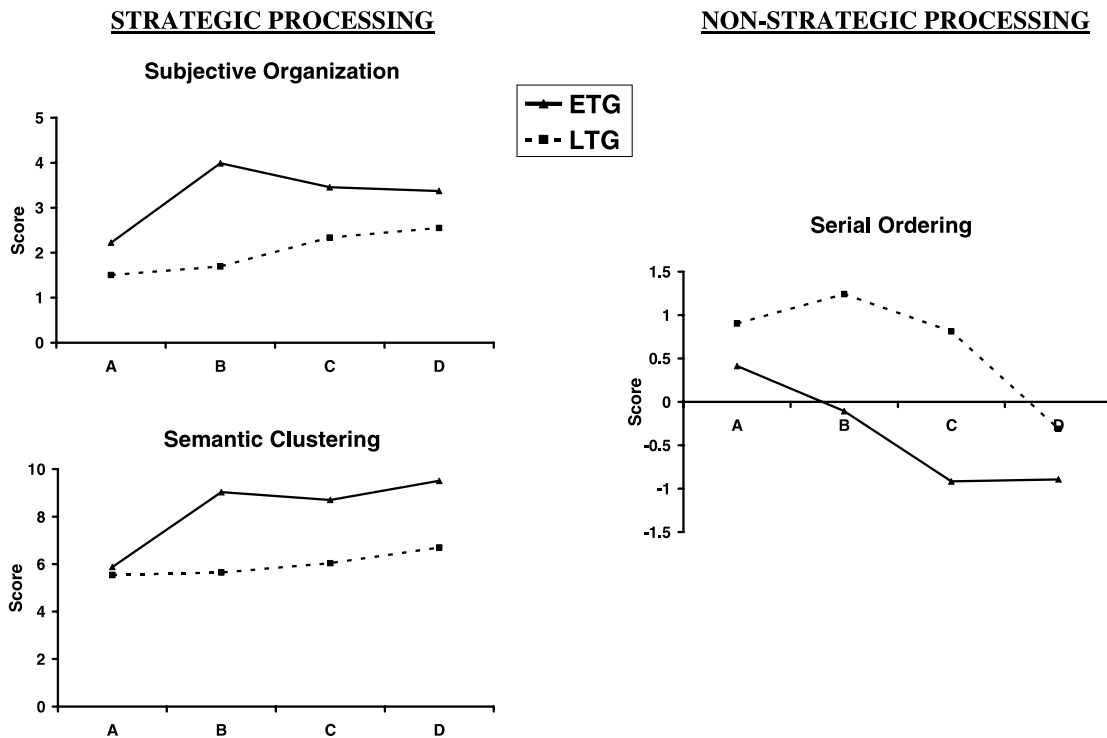


Fig. 3. Strategic and nonstrategic performance by the Early (ETG) and Late (LTG) Training Groups at all assessments (A, B, C, D) of the Hopkins Verbal Learning Test-Revised (HVLTR). *Subjective organization* was measured by Pair Frequency Analysis (Sternberg & Tulving, 1977). This measure, which adjusts for the number of words recalled, tabulates the number of word pairs recalled together from one trial to the next. *Semantic clustering* was measured as the number of consecutively recalled words from the same semantic category. *Serial ordering* is the number of words recalled in the same order as presented. A proportional measure was obtained by dividing the number of serial order clusters by the theoretically maximal number of order clusters, which in turn depends on the total number of words recalled.

Long-term effects of training (Assessment D)

Within-group comparisons of scores between Assessments A and C revealed significant improvement in the ETG on words recalled at the 36-s delay ($F_{1,25} = 8.59, p = .007$);

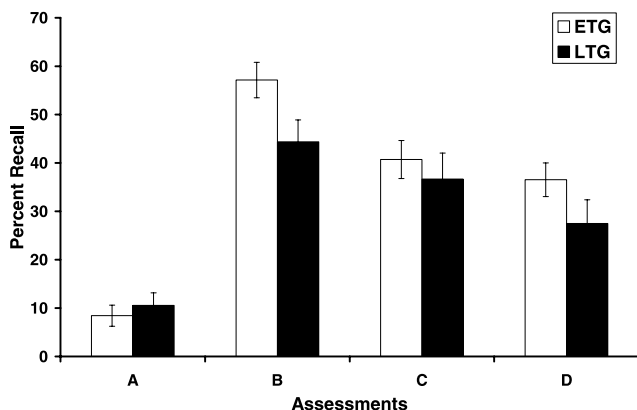


Fig. 4. Percent delayed recall of Level 3 information by the Early (ETG) and Late (LTG) Training Groups at all assessments of the Logical Stories Test. Error bars represent the standard error of the mean.

however, this difference was not significant between Assessments A and D. Improvement on the strategy measure was in the expected direction, but was not statistically significant.

Summary

At Assessment C, within-group comparisons indicated that the LTG improved relative to Assessment B on both secondary memory and strategy measures. Similar comparisons provided some evidence of relatively long-term benefits in secondary memory in the ETG.

HVLT

There were no significant between-group differences in words recalled from primary memory in immediate recall or in delayed recognition memory at any of the testing sessions, using ANCOVA. (ANOVA conducted on the primary memory measure did not reveal significant group differences at any of the assessments.)

Training effects on ETG (Assessments A to B)

As mentioned in the statistical analysis section, at Assessment A, 7 participants were allowed to respond in a written

Table 1. Means (and *SD*) for main measures at the four serial assessments

	Assessment A		Assessment B		Assessment C		Assessment D	
	ETG (<i>n</i> = 29)	LTG (<i>n</i> = 20)	ETG (<i>n</i> = 29)	LTG (<i>n</i> = 20)	ETG (<i>n</i> = 26)	LTG (<i>n</i> = 15)	ETG (<i>n</i> = 28)	LTG (<i>n</i> = 15)
Alpha Span								
<i>Working Memory</i>								
Total Score	5.38 (1.66)	5.40 (2.01)	5.66 (1.20)	5.50 (1.99)	6.04 (1.25)	5.93 (1.33)	5.12 (1.24) ^b	5.36 (1.39) ^a
Partial Score	23.34 (8.99)	23.45 (11.82)	24.28 (7.51)	24.45 (11.99)	25.12 (8.77)	24.40 (8.19)	21.27 (8.00) ^b	22.86 (10.06) ^a
Brown Peterson								
<i>Secondary Memory</i>								
36-s Delay	3.59 (2.28)	3.75 (2.51)	4.00 (2.74)	3.40 (2.54)	4.77 (2.63)	4.00 (3.12)	4.48 (2.68) ^c	3.29 (2.58) ^a
<i>Primary Memory</i>								
PM 0-s Delay	.86 (.18)	.91 (.18)	.90 (.16)	.82 (.17)	.87 (.21)	.89 (.15)	.86 (.24) ^c	.91 (.14) ^a
PM 3-s Delay	.64 (.30)	.71 (.20)	.68 (.20)	.65 (.23)	.64 (.27)	.59 (.26)	.68 (.26) ^c	.74 (.19) ^a
PM 18-s Delay	.15 (.16)	.23 (.21)	.13 (.13)	.21 (.20)	.13 (.22)	.19 (.27)	.19 (.23) ^c	.21 (.27) ^a
Performance	1.06 (.24)	1.10 (.28)	1.13 (.28)	1.11 (.27)	1.20 (.25)	1.19 (.31)	1.15 (.28) ^c	1.13 (.28) ^a
Strategy	−.46 (.21)	−.48 (.24)	−.46 (.26)	−.54 (.25)	−.41 (.24)	−.53 (.29)	−.41 (.28) ^c	−.58 (.22) ^a
HVLT								
<i>Delay</i>								
Total Recall	9.05 (2.34) ^e	8.37 (3.17) ^a	9.41 (2.51)	8.25 (2.51)	9.12 (2.78)	8.60 (3.18)	9.68 (2.36)	8.00 (4.17)
Recognition	10.45 (1.53) ^e	10.32 (2.36) ^a	10.03 (1.78)	9.80 (2.40)	10.62 (1.20)	10.67 (1.35)	10.86 (1.46)	10.40 (1.84)
<i>Immediate</i>								
Primary Memory Recall	6.36 (2.74) ^e	6.37 (2.14) ^a	6.24 (2.61)	6.85 (2.32)	7.42 (1.77)	6.53 (2.29)	6.18 (1.98)	6.53 (1.92)
Secondary Memory Recall	18.41 (5.42) ^e	18.00 (5.57) ^a	20.79 (6.03)	17.25 (5.79)	19.50 (5.78)	19.53 (6.13)	21.68 (5.31)	18.60 (7.75)
Serial Ordering	.41 (2.01) ^e	.90 (1.91) ^a	−.11 (1.79)	1.24 (2.03)	−.92 (1.77)	.81 (2.89)	−.89 (1.88)	−.31 (2.01)
Subjective Organization	2.22 (2.19) ^e	1.51 (1.55) ^a	3.99 (2.68)	1.70 (2.60)	3.45 (2.65)	2.34 (2.93)	3.37 (2.23)	2.55 (3.64)
Semantic Clustering	5.88 (5.52) ^e	5.54 (5.09) ^a	9.03 (5.59)	5.65 (5.06)	8.71 (5.73)	6.04 (6.54)	9.51 (5.77)	6.70 (6.43)
Logical Stories								
Level 3 Immediate Recall	13.79 (10.14)	17.22 (19.24)	60.34 (18.32)	46.05 (22.07) ^a	49.77 (23.00)	43.93 (28.93) ^a	48.21 (16.70) ^d	30.50 (18.76) ^a
Level 3 Delayed Recall	8.43 (9.69)	10.56 (14.18)	56.03 (20.49)	46.05 (26.37) ^a	43.04 (22.44)	32.36 (27.08) ^a	39.29 (18.13) ^d	22.57 (14.87) ^a
Total Immediate Recall	51.10 (11.87)	44.85 (19.41)	66.81 (11.81)	56.14 (15.98) ^a	67.15 (16.41)	54.71 (26.28) ^a	66.47 (10.26) ^d	49.71 (14.44) ^a
Total Delayed Recall	44.31 (11.83)	38.79 (18.09)	63.22 (12.75)	52.85 (18.58) ^a	61.23 (20.44)	49.12 (24.07) ^a	59.88 (11.22) ^d	46.45 (14.35) ^a

Note. ETG = Early Training Group; LTG = Late Training Group; HVLT = Hopkins Verbal Learning Test – Revised.

^aOne missing value relative to sample size identified at the top of the column.

^bTwo missing values.

^cThree missing values.

^dFour missing values.

^eSeven missing values.

format which they were able to use as a type of memory prosthetic resulting in improved performance. We opted to compare HVLTL scores for the ETG and LTG at Assessment B using analysis of variance without covarying Assessment A and drastically reducing sample size. An impact of rehabilitation in immediate memory testing was a substantial and significant group difference in the number of words recalled from secondary memory ($F_{1,47} = 4.22, p = .05, \eta^2 = .082$). At Assessment B, group differences on the other recall measures (immediate total recall and delayed recall) were all in the predicted direction, although differences just failed to be significant (all p 's $> .05 < .08$). Immediate recall scores are shown in Figure 2.

Turning to measures of strategy use, the first to be considered is *serial order*. Given that the 12-word HVLTL lists contained three semantic categories of 4 words each, an efficient strategy is to encode and retrieve the words in terms of these categories. It is, therefore, likely that good retrieval performance will be associated with a shift away from reproducing the randomized order of presentation (that is, a decline in serial ordering) and a shift toward retrieving same-category words together (that is, a rise in semantic clustering). If such strategic awareness is a consequence of memory training, we expect these shifts to occur predominantly between Assessments A and B for the ETG, and between Assessments B and C for the LTG.

Serial order scores for participants who completed all four batteries are shown in Figure 3. Serial order clustering declined in the ETG at Assessment B, relative to Assessment A, while rising slightly, but not significantly, in the LTG. This finding resulted in a significant group difference on this measure, $Z = -2.74, p = .01$ (nonparametric analysis was used because the distribution of scores was not normal), favoring the ETG. Conversely, and as expected, the use of semantic clustering and subjective organization strategies increased in the ETG, relative to the LTG (Figure 3). ANOVA, applied to these scores, indicated significant group differences in semantic clustering ($F_{1,47} = 4.68, p = .04, \eta^2 = .091$), and subjective organization ($F_{1,47} = 8.90, p = .005, \eta^2 = .159$).

To a limited extent, within-group comparisons across Assessments A to B provided support for greater use of effective strategies by the ETG at Assessment B. This group showed a significant increase in semantic clustering ($F_{1,20} = 4.39, p = .05$), and a numerically smaller decrease in serial ordering ($F_{1,20} = 3.72, p = .07$). Between Assessments A and B, there were no significant changes within the LTG on any of the strategic measures.

Training effects on LTG (Assessments B to C)

At Assessment C, ANCOVA did not show significant differences between the ETG and the LTG on any of the memory measures, or on measures of subjective organization or semantic clustering. There was a significant group difference on the serial order measure at Assessment C ($Z = -2.12, p = .03$; nonparametric analysis was performed at

Assessment C between groups because the scores were not distributed normally), indicating that the LTG made greater use of this inefficient strategy than the ETG (Figure 3). Within-group comparisons strengthened the suggestion that the LTG had benefited from rehabilitation training. Relative to Assessment B, at immediate memory testing, the LTG improved significantly in terms of total recall ($F_{1,14} = 7.88, p = .01$), and recall from secondary memory ($F_{1,14} = 4.68, p = .05$), while ETG showed negligible change for total recall ($F_{1,23} = .70, p = .41$) and nominal change in secondary memory recall ($F_{1,23} = 3.60, p = .07$).

Comparisons between Assessments A and C yielded no significant differences for the LTG, but the ETG showed significant improvements in immediate recall for total correct words ($F_{1,18} = 5.92, p = .03$). For this comparison, the ETG also exhibited improvements in semantic clustering ($F_{1,18} = 8.10, p = .01$), a marginally significant improvement in subjective organization ($F_{1,18} = 3.95, p = .06$), and a significant decline in serial ordering ($F_{1,18} = 10.73, p = .004$; see Figure 3).

Long-term effects of training (Assessment D)

There were no significant differences between the ETG and the LTG at Assessment D. However, within-group comparisons indicated long-term benefits of rehabilitation training in the ETG. For example, relative to baseline testing (Assessment A), the ETG improved in total words recalled ($F_{1,16} = 18.21, p < .001$), and in words recalled from secondary memory ($F_{1,16} = 10.46, p = .005$). The ETG also exhibited an increased use of semantic clustering ($F_{1,16} = 9.57, p = .007$), a marginally significant increase in subjective organization ($F_{1,16} = 4.12, p = .06$), and a decreased use of serial ordering ($F_{1,16} = 13.51, p = .002$). Similar comparisons failed to yield significant changes in the LTG.

Summary

The ETG and LTG showed immediate benefits of training on the HVLTL-R test, both in terms of number of words recalled and efficient use of strategies. The ETG, but not the LTG, continued to show improvement in performance on these measures at Assessment D, indicating that the effects of training for this group were long-lasting.

Logical Stories Test

To simplify presentation, of the three levels of analysis only the results of Level 3 recall (specific details) are presented here, along with measures of total recall.

Training effects on ETG (Assessments A to B)

In line with our hypothesis that the beneficial effects of training would be especially apparent in strategic functioning, large effects were observed in recall of Level 3 information. In Level 3, there were significant group differences at Assessment B in immediate recall ($F_{1,45} = 9.60, p =$

.003, $\eta^2 = .176$), and in delayed recall ($F_{1,45} = 4.77$, $p = .03$, $\eta^2 = .096$; the data for Delayed Recall of Level 3 information are provided in Figure 4). These differences favored the ETG and reflect the effect of rehabilitation training.

Between Assessments A and B, both groups showed significant improvement in *total* immediate recall (ETG: $F_{1,28} = 28.80$, $p < .001$; LTG: $F_{1,18} = 11.79$, $p = .003$), and *total* delayed recall (ETG: $F_{1,28} = 55.16$, $p < .001$; LTG: $F_{1,18} = 18.01$, $p < .001$). Despite the unexpected improvement in both groups at Assessment B, the ETG performed significantly better than the LTG in immediate recall ($F_{1,45} = 4.52$; $p = .04$). A nonsignificant difference was found to be in the same direction for delayed recall ($F_{1,45} = 3.17$; $p = .08$).

The overall pattern of results at Assessment B indicates benefits of rehabilitation training for the ETG in Level 3 information (recall of specific detail), but no specific benefit for total recall. The equivalent increase in total recall scores for both groups may reflect a general practice effect or different levels of difficulty of the stories at the two tests.

Training effects on LTG (Assessments B to C)

There was no evidence of improvement in the LTG, relative to the ETG, on any of the measures at Assessment C.

Long-term effects of training (Assessment D)

At Assessment D, the ETG recalled slightly more Level 3 information than the LTG in immediate recall, with the group difference being marginally significant ($F_{1,34} = 3.65$, $p = .06$). There was no significant difference on this measure at delayed recall ($F_{1,34} = 2.03$, $p = .16$). At this session, the ETG performed significantly better than the LTG in terms of *total* words recalled at immediate ($F_{1,34} = 12.19$, $p = .001$) and delayed ($F_{1,34} = 5.92$, $p = .02$) recall.

Within-group comparisons confirmed that there were no significant differences in the scores of either group between Assessments C and D. However, both groups exhibited significant improvements from Assessments A to D (for Level 3 immediate recall: ETG, $F_{1,23} = 140.64$, $p < .001$; LTG, $F_{1,13} = 11.23$, $p = .005$; and delayed recall: ETG, $F_{1,23} = 83.82$, $p < .001$; LTG, $F_{1,13} = 25.53$, $p < .001$. Total immediate recall: ETG, $F_{1,23} = 34.33$, $p < .001$; LTG, $F_{1,13} = 5.03$, $p = .04$; Total delayed recall: ETG, $F_{1,23} = 25.27$, $p < .001$; LTG, $F_{1,13} = 23.86$, $p < .001$). Any conclusion about long-term benefits of training on this measure must be qualified by the possible influence of practice effects.

Summary

In the Logical Stories Test, the ETG showed some benefits of training in immediate and delayed recall. Relative to baseline, both groups also improved in Level 3 information recalled, an indication that they were using more efficient strategies. The ETG retained training-derived benefits at Assessment D; this finding is especially evident in the total

recall scores. LTG performance is also higher than baseline on all measures, but interpretation of these results must be qualified in view of apparent practice effects and possible differences in story difficulty across test sessions.

DISCUSSION

Before conducting the study, our assumption was that training would have its greatest effects on those types of memory that involve more controlled and strategic processing, for example, working memory and free recall (e.g., Craik & Jennings, 1992). Contrary to expectations, our measure of working memory, Alpha Span, showed no effects of training. Speculatively, the kinds of mental operations involved in holding and re-ordering word sequences are not really “strategies” but rather operations that would become more efficient as the result of long practice in the same way as mental arithmetic operations improve with practice.

Primary memory, in the sense of Waugh and Norman (1965), was measured in the Brown-Peterson and HVLTR tests, and showed no meaningful changes as a function of training. This null result makes sense in that primary memory is not sensitive to strategy manipulations; it is also in line with findings that this type of memory is affected very little by aging (Craik, 1968; Craik & Jennings, 1992). Recognition memory also shows relatively slight effects of aging (Craik & Jennings, 1992; Craik, 1983) and was also unaffected by training.

On the other hand, recall from secondary memory did show beneficial effects of training. This is an important result given that secondary memory is the type of episodic memory in constant use in our daily lives. In the HVLTR test, the ETG performed significantly better than the LTG at Assessment B on total words recalled. This group difference disappeared at Assessment C, due largely to improved performance in the LTG participants following their training. The benefits of training were long lasting in the ETG as they continued to show improved performance at Assessment D, relative to baseline at Assessment A. By comparison, the LTG did not lose gains between training and Assessment D testing, but there was no evidence of continued improvement beyond Assessment C. Of note, within the HVLTR data, there were changes in the use of strategies, especially by the ETG. The general pattern, which characterized the ETG’s performance at all posttraining assessment sessions, was increased use of subjective organization and semantic clustering strategies, and a decline in the use of less efficient serial ordering. As noted, because Assessment A was not covaried, it is possible that the visible baseline difference in strategy performance at Assessment A contributed to the statistical differences; nevertheless, the pattern of increased use of efficient and decreased use of inefficient strategies would remain.

The results of the Logical Stories Test are somewhat ambiguous because of practice effects seen in both groups. It is also possible that the stories varied in difficulty, although significant variation is considered unlikely. The stories were

taken from Dixon et al. (1989) who had standardized the collection and reported them to be comparable in terms of content, length, and difficulty. Nonetheless, recall of Level 3 information improved significantly following training in both groups, although the effect was consistently greater in the ETG. Given the strategic aspect of recalling such information, this finding provides further validation for targeting strategic processes in rehabilitation training.

Although both groups responded positively to training, as in other domains (e.g., psychosocial) the ETG exhibited more consistent and long-lasting benefits than the LTG. This discrepancy was addressed in some detail in the Psychosocial paper (Winocur et al., 2007). To reiterate, we believe that the most likely explanation lies in the groups' responses to the initial orientation and the failure of LTG participants to appreciate that there would be a delay before the start of their training program. We suspect that this factor translated into a negative reaction of which the participants themselves may not have been aware.

The current results should be interpreted cautiously because of the number of measures, the potential power issue due to the unavoidable loss of participants at Assessments C and D, and the small size of the effects. On the other hand, it is encouraging that the results consistently were in the expected direction. It is enlightening that the effects on memory (e.g., recall of word lists and stories) were seen primarily on tests that involve the greatest degree of strategic control during encoding and retrieval (e.g., Cavallini et al., 2002). Conversely, primary memory and recognition tests, which are "well supported" by ongoing processing and by the test environment (Craik, 1983), and not affected by aging (Craik & Jennings, 1992), did not benefit from rehabilitation training.

This research adds to the literature that training can improve encoding and retrieval processes in the elderly, leading to better memory functioning (Glisky & Glisky, 1999) over extended periods of time (Stigsdotter Neely & Bäckman, 1993). Although we assume that the effects were due largely to the specific memory training module, the extent to which this is so remains uncertain. It is possible that other modules, a combination of modules, or simply involvement in the study, provided the effects. As well, our data suggested that the training had some effect on real life situations (Levine et al., 2007), and appeared to generalize to a cognitive domain not specifically targeted in training (i.e., language; see Winocur et al., 2007). Future research is needed to assess the generalizability to practical, everyday aspects of memory.

ACKNOWLEDGMENTS

The authors acknowledge the valuable contributions of the other co-investigators on this project: Drs. M. Alexander, S. Black, D. Dawson, B. Levine, and I.H. Robertson. As well, the outstanding support provided by Maureen Downey-Lamb, Louise Fahy, Marina Mandic, and Tara McHugh is gratefully acknowledged. This study, and the experimental trial of which it is a part, were supported by

the JSF McDonnell Foundation. D.T. Stuss holds the Reva James Leeds Chair in Neuroscience and Research Leadership at Baycrest and the University of Toronto. The information reported in this manuscript and the manuscript itself are new and original. The manuscript is not under review by any other journal and has never been published either electronically or in print. There are no financial or other relationships that could be interpreted as a conflict of interest affecting this manuscript.

REFERENCES

- Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8) (pp. 47–90). New York: Academic Press.
- Baddeley, A.D. & Warrington, E.K. (1970). Amnesia and distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 176–189.
- Ball, K., Berch, D.B., Helmers, K.F., Jobe, J.B., Leveck, J.D., Marsiske, M., Morris, J.N., Rebok, G.W., Smith, D.M., Tennstedt, S.L., Unverzagt, F.W., & Willis, S.L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, 233, 2271–2281.
- Benedict, R.H.B., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test – Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist*, 12, 43–55.
- Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, 5, 125–142.
- Cavallini, E., Pagnin, A., & Vecchi, T. (2002). The rehabilitation of memory in old age: Effects of mnemonics and metacognition in strategic training. *Clinical Gerontologist*, 26, 125–141.
- Craik, F.I.M. (1968). Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*, 7, 996–1004.
- Craik, F.I.M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, Series B*, 302, 341–359.
- Craik, F.I.M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities, mechanisms, and performance* (pp. 409–422). Amsterdam: North-Holland.
- Craik, F.I.M. & Jennings, J.J. (1992). Human memory. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.
- Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Craik, F.I.M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294.
- Dixon, R.A., Hultsch, D.F., & Hertzog, C. (1989). *A manual of three-tiered structurally equivalent texts for use in aging research*. (CRGCA Technical Report No. 2). Victoria, BC: University of Victoria.
- Engle, R.W. & Kane, M.J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* 44 (pp. 145–199). New York: Elsevier.
- Floden, D., Stuss, D.T., & Craik, F.I.M. (2000). Age differences in performance on two versions of the Brown-Peterson task. *Aging, Neuropsychology, and Cognition*, 7, 245–259.

- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Minimal state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.
- Glisky, E.L. & Glisky, M.L. (1999). Memory rehabilitation in the elderly. In D.T. Stuss, G. Winocur, & I.H. Robertson (Eds.), *Cognitive rehabilitation* (pp. 347–361). Cambridge: Cambridge University Press.
- Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L.L., Jennings, J.M., & Hay, J.F. (1996). Dissociating automatic and consciously controlled processes: Implications for diagnosis and rehabilitation of memory deficits. In D.J. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M.K. Johnson (Eds.), *Basic and applied memory research: Theory in context* (pp. 161–193). Mahwah, NJ: Lawrence Erlbaum.
- Jennings, J.M. & Jacoby, L.L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*, 283–293.
- Levine, B., Stuss, D.T., Winocur, G., Binns, M.A., Fahy, L., Mandic, M., Bridges, K., & Robertson, I.H. (2007, this issue). Cognitive rehabilitation in the elderly: Effects on strategic behavior in relation to goal management. *Journal of the International Neuropsychological Society*, *13*, 143–152.
- Scogin, F., Storandt, M., & Lott, L. (1985). Memory-skills training, memory complaints, and depression in older adults. *Journal of Gerontology*, *40*, 562–568.
- Sheikh, J.I., Hill, R.D., & Yesavage, J.A. (1986). Long-term efficacy of cognitive training for age-associated memory impairment: A six-month follow-up study. *Developmental Neuropsychology*, *2*, 413–421.
- Sternberg, R.J. & Tulving, E. (1977). The measurement of subjective organization in free recall. *Psychological Bulletin*, *84*, 539–556.
- Stigsdotter Neely, A. & Bäckman, L. (1993). Long-term maintenance of gains from memory training in older adults: Two 3 1/2 year follow-up studies. *Journal of Gerontology: Psychological Sciences*, *48*, 233–237.
- Stuss, D.T., Robertson, I.H., Craik, F.I.M., Levine, B., Alexander, M.P., Black, S., Dawson, D., Binns, M.A., Palmer, H., Downey-Lamb, M., & Winocur, G. (2007, this issue). Cognitive Rehabilitation in the elderly: A randomized trial to evaluate a new protocol. *Journal of the International Neuropsychological Society*, *13*, 120–131.
- Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *5*, 193–197.
- Tulving, E. & Colotla, V.A. (1970). Free recall of trilingual lists. *Cognitive Psychology*, *1*, 86–98.
- Waugh, N.C. & Norman, D.A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.
- Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale-Revised, manual*. San Diego: The Psychological Corporation.
- Winocur, G., Palmer, H., Dawson, D., Binns, M.A., Bridges, K., & Stuss, D.T. (2007, this issue). Cognitive rehabilitation in the elderly: An evaluation of psychosocial factors. *Journal of the International Neuropsychological Society*, *13*, 153–165.