

Episodic memory change in late adulthood: Generalizability across samples and performance indices

ROGER A. DIXON

University of Alberta, Edmonton, Alberta, Canada

ÅKE WAHLIN

Stockholm University, Stockholm, Sweden

SCOTT B. MAITLAND

University of Guelph, Guelph, Ontario, Canada

DAVID F. HULTSCH

University of Victoria, Victoria, British Columbia, Canada

CHRISTOPHER HERTZOG

Georgia Institute of Technology, Atlanta, Georgia

and

LARS BÄCKMAN

Karolinska Institute, Stockholm, Sweden

Younger adults recall more information from episodic memory tasks than do older adults. Because longitudinal studies are rare and often incompatible, the extent of actual late-life memory change is not well established. We assemble two different longitudinal samples of normal older adults, each of which is tested twice at a 3-year interval, using a large battery of episodic memory indicators. Together, two-wave data from both the Victoria Longitudinal Study in Canada ($n = 400$) and the Kungsholmen Project in Sweden ($n = 168$) cover a 40-year span of adulthood, ranging from 54 to 94 years of age. Principal memory tasks include categorizable word lists, story recall, and random word lists, as well as indicators of cognitive support. Overall, an examination of performance on sets of common and complementary episodic tasks reveals that, for both samples, actual 3-year changes are modest and that, when decline occurs, it is gradual. The exception—greater decline for more supported tasks—suggests that these may be especially sensitive to late-life changes.

Episodic memory is one of the most frequently examined and age-sensitive aspects of cognitive performance in adults. Across a variety of episodic tasks, younger adults typically remember substantially more target information

than do older adults (e.g., Bäckman, Small, Wahlin, & Larsson, 2000; Zacks, Hasher, & Li, 2000). Cross-sectional comparisons of extreme groups (e.g., college students vs. older adults) dominate the available evidence on aging and episodic memory. Naturally, such between-group comparisons cannot directly support inferences regarding within-person, aging-related changes. In contrast, the less frequently used longitudinal designs provide the most direct evidence pertaining to actual change (e.g., Schaie & Hofer, 2001). Moreover, recently developed techniques for addressing well-known methodological challenges of change-sensitive designs have been applied successfully to longitudinal cognitive aging research (e.g., Hultsch, Hertzog, Dixon, & Small, 1998; Schaie, 1996).

Although the results of both cross-sectional and longitudinal studies lead to inferences of aging-related decline in episodic memory performance, inconsistencies are apparent within and across the two sets of designs (Hertzog, 2004). Specifically, although aging-related

The Victoria Longitudinal Study is supported by Grant AG08235 from the National Institute on Aging to R.A.D., who acknowledges additional support from the Canada Research Chairs program. The Kungsholmen Project is supported by grants from the Swedish Council for Research in the Humanities and Social Sciences and the Swedish Council for Social Research to L.B. We appreciate the helpful comments of Anna-Lisa Cohen, Cindy de Frias, Lars-Göran Nilsson, and anonymous reviewers on earlier drafts of this article. The 3-year cross-sectional versus longitudinal comparisons in the discussion were suggested by Peter Graf. Correspondence should be addressed to R. A. Dixon, Department of Psychology, P217 Biological Sciences Building, University of Alberta, Edmonton, AB, T6G 2E9 Canada (e-mail: rdixon@ualberta.ca).

Note—This article was reviewed and accepted by the previous editorial team, under Morton Ann Gernsbacher.

longitudinal decline has been regularly observed, the magnitude and rate of decline are rarely as great as would be expected from the results of most extreme-groups cross-sectional comparisons. Notably, however, there is also some inconsistency in the results of longitudinal studies. This inconsistency in the magnitude and rate of observed longitudinal change in memory performance has been related to important methodological factors, including (1) differential longitudinal retest intervals (e.g., Zelinski & Burnight, 1997), (2) older adult samples varying widely in demographic, age, and other memory-relevant characteristics (e.g., Bäckman, Jonsson Laukka, Wahlin, Small, & Fratiglioni, 2002; Giambra, Arenberg, Zonderman, Kawas, & Costa, 1995), and (3) the use of single or unique memory indicators (e.g., Small, Dixon, Hultsch, & Hertzog, 1999). The present research plan pivoted on the opportunity to evaluate aspects of these factors. Specifically, we examined actual change in multiple episodic memory measures in two different longitudinal samples, which together cover a 40-year band of older adults.

We will elaborate on the importance of these methodological factors to the present study. First, regarding the period between retest occasions, longitudinal studies on episodic memory with intervals ranging up to 5 years may report relatively modest aging decline (e.g., Small et al., 1999; Zelinski, Gilewski, & Schaie, 1993). In contrast, studies with longer retest intervals (e.g., up to 16 years) typically report significant declines in level of performance (e.g., Colsher & Wallace, 1991; Giambra et al., 1995; McCarty, Siegler, & Logue, 1982; Taylor, Miller, & Tinklenberg, 1992; Zelinski & Burnight, 1997). Overall, however, the extent of decline is relatively small. In one study, 6-year declines ranged from 2% to 4% of original performance (Small et al., 1999). Thus, although extreme-groups cross-sectional comparisons reveal substantial adult age differences in episodic memory performance, the actual rate of memory changes in late life may be more gradual than precipitous. This possibility was evaluated in the present study in two ways: (1) A 3-year interval is used, so even modest changes can be charted, and (2) 3-year change data from two longitudinal studies are included.

Second, age and other participant characteristics may affect results. Typically, cross-sectional and longitudinal studies of memory and aging recruit volunteer samples of convenience, usually including one group of older adults. Despite stereotypes to the contrary, older adults are not a homogeneous group, and any single convenience sample may not be representative. Moreover, long-term cognitive aging processes may vary by basic sample characteristics, such as age, health, education, biological condition, and background (Bäckman et al., 2000; MacDonald, Dixon, Cohen, & Hazlett, 2004; Wahlin, 2004). In longitudinal research, convenience samples may be especially vulnerable to initial selection and attrition effects. Such effects may bias the sample toward a more able (e.g., in educational experience, younger age) or less advantaged (e.g., in health or neurological conditions) group, which may be associated with either an over- or an underestimation of

cognitive maintenance. Fortunately, there are several ways of examining the extent to which selection factors are involved in the determination of longitudinal trends observed in specific samples (e.g., Hultsch et al., 1998; Schaie, 1996; Sliwinski, Lipton, Buschke, & Stewart, 1996). One promising, but as yet unexplored, way is to compare trends across independent samples from different longitudinal studies (Schaie & Hofer, 2001). Because longitudinal studies are expensive and uncommon and often contain nonoverlapping methodological features, they are rarely compared systematically. Nevertheless, such comparisons are valuable in understanding the extent to which generalizability across selection factors may be permitted. Therefore, in the present article, we assemble data from two separate but methodologically complementary longitudinal studies, one with a large convenience sample and the other with a population-based sample. Moreover, because we follow adults who range in age from their 50s to their 90s, we can determine whether rates of change vary across a substantial portion of the adult life span.

The third methodological factor reflects the diversity of episodic memory indicators used in cognitive aging research: Comparing performance and change patterns is complicated when only single or unique sets of tasks are used. Perhaps the most typical indicator has been total amount of information recalled (e.g., free recall of words or texts). Other potential indicators include measures of clustering or output structure, free recall performance under various conditions (such as presentation rate), and performance with nonverbal materials, as well as cued recall and recognition. Multiple indicators of episodic memory performance are available in both of the present longitudinal studies; our measures range from the similar to the complementary. In both studies, organizable word lists were used, and estimates of correct free recall, errors, and clustering performance were derived. For free recall, previous longitudinal research would predict only modest declines, although perhaps more so for very old adults and for population-based samples (e.g., Small et al., 1999; Zelinski & Burnight, 1997). The structural organization of the output, or clustering performance (e.g., Roenker, Thompson, & Brown, 1971), is more complicated, since the only previous longitudinal study showed general preservation in a convenience sample (Small et al., 1999), but a corresponding cross-sectional study reported that younger adults clustered their recall to a greater extent than did older adults (Bäckman & Larsson, 1992). Finally, in the present article, unique longitudinal data are available on performance on story recall, random word list recall, cued and recognition measures, and derived indicators of primary and secondary memory performance. In sum, a broad sampling of the domain of episodic memory is available in this report—with all measures collected on two occasions of measurement.

Through the use of a longitudinal design with two independent samples, the present study contributes to the growing literature on aging effects in memory perfor-

mance. Specifically, we included a volunteer sample of 400 older adults from the Canadian Victoria Longitudinal Study (VLS) and a population-based sample of 168 older adults from the Swedish Kungsholmen Project (KP). The VLS and the KP feature matching, complementary, and unique characteristics. Importantly, both studies began in the 1980s, both had 3-year intervals between testing occasions, and both followed a subset of older adults. Together, two-wave change patterns are available for a combined age range of 55–95 years. Four different 10-year age groups were derived in these samples, making possible both cross-sectional and longitudinal comparisons. Moreover, both samples include the age range of 75–84 year olds, a period in which some observers have argued accelerated cognitive decline may occur (e.g., Giambra et al., 1995; Small et al., 1999). The present samples have been selected to represent normal aging, diminishing the potential influences of health-related and neurodegenerative decline. Finally, the two studies contain multiple indicators of episodic memory, both parallel and unique measures.

METHOD

Participants

Victoria Longitudinal Study. The participants were 400 adults from VLS Sample 2, who completed the two waves of data collection, approximately 3 years apart ($M = 3.22$ years, $SD = 0.17$). Of the 530 adults at baseline, 10 were excluded from this study because they did not fit the present age requirements. Attrition of the 120 nonreturners occurred for typical reasons, most prominently personal illness (36.1%) or busy and insufficient interest (22.1%). Attrition analyses compared 400 returners with 120 nonreturners on first-wave demographic indicators and memory performance. As was expected, the returners were, on average, 3 years younger than the nonreturners, and the returners recalled more information than did the nonreturners ($ps < .05$).

For purposes of comparing longitudinal gradients with the KP, the VLS sample was divided into three age groups of approximately 10 years. The young-old age group (YO; $n = 148$) ranged from 54 to 64 years ($M = 60.31$, $SD = 2.92$). The mid-old age group (MO; $n = 181$) ranged from 65 to 74 years ($M = 69.23$, $SD = 2.84$). The old-old age group (OO; $n = 71$) ranged from 75 to 84 years ($M = 78.73$, $SD = 2.60$). Women made up 66% of the YO group, 62% of the MO group, and 72% of the OO group. The mean numbers of years of education for the YO group ($M = 15.33$ years, $SD = 3.13$), the MO group ($M = 15.12$ years, $SD = 2.86$), and the OO group ($M = 14.00$ years, $SD = 3.23$) are relatively high but are similar to those for an earlier VLS sample (Hultsch et al., 1998). The participants generally rated their health as being good to very good on a single-item 5-point scale, in which 0 indicated *very good health* ($M_{YO} = 0.70$, $M_{MO} = 0.67$, $M_{OO} = 0.75$). In order to compare their health with that in the KP sample, we selected a set of 14 conditions for which data were collected by both studies (e.g., arthritis, emphysema, hardening of arteries, high blood pressure, heart trouble, stroke, diabetes). Generally low proportions of the VLS age groups reported experiencing these conditions ($M_{YO} = 6.61\%$, $SD = 7.22$; $M_{MO} = 7.77\%$, $SD = 8.08$; $M_{OO} = 10.16\%$, $SD = 7.60$). A 3 (age) \times 2 (gender) multivariate analysis of variance (MANOVA) on the background characteristics revealed significant effects for both factors on years of education and proportion of common illnesses. The OO participants had fewer years of education than did the two younger groups, and also reported more common illness conditions

than did the YO group. Men ($M = 15.57$, $SD = 3.38$) had more years of education than did women ($M = 14.58$, $SD = 3.04$). On the single-item indicator of health, women ($M = 1.64$, $SD = 0.67$) perceived their health as better than did men ($M = 1.78$, $SD = 0.67$).

Kungsholmen Project. The present KP sample was taken from a population-based recruitment of all adults over age 75 living in the Kungsholmen community of Stockholm, Sweden. The assessment included complete medical examinations, psychiatric evaluations, neurological testing, cognitive testing, and social and family interviews (see Fratiglioni, Viitanen, Bäckman, Sandman, & Winblad, 1992). The participants were 168 adults initially ranging in age from 75 to 94 years, who were tested twice approximately 3 years apart ($M = 3.10$ years, $SD = 0.73$). Of the 368 healthy individuals who participated in the baseline cognitive testing, attrition occurred for 147. The most prominent reasons for attrition were death (50%) and refusal to participate further (45%). In addition, 53 baseline participants who received a dementia or psychiatric diagnosis at the second occasion were excluded from this study. Parallel attrition analyses were conducted, in which 168 returners were compared with the 147 nonreturners. As was expected, the returners were, on average, 3 years younger than the nonreturners and had engaged in about 1 year more of formal schooling. Furthermore, returners recalled more information than did nonreturners. The gender distributions were equivalent in the two groups.

The KP sample was divided into two 10-year age groups, such that one overlapped with the VLS–OO group and the other formed an even older group. The KP old-old group ranged from 75 to 84 years ($M_{OO} = 79.88$, $SD = 2.93$), and the very old group ranged from 85 to 94 years ($M_{VO} = 88.05$ years, $SD = 2.48$). Regarding formal schooling, the KP–OO group reported $M = 9.07$ years ($SD = 3.01$), and the KP–VO group reported $M = 9.13$ years ($SD = 2.95$). These levels are not significantly different from one another, and are representative of these cohorts in Stockholm (Statistiska Centralbyrån, 1997). As with the VLS, women predominated in the two age groups, making up 74% of the KP–OO and 86% of the KP–VO groups. The common indicators of actual health conditions revealed that the KP groups were only slightly less healthy than the VLS sample. The KP–OO group reported $M = 12.36\%$ ($SD = 8.85$), and the KP–VO group reported $M = 12.07\%$ ($SD = 9.00$) of the 14 common health conditions.

Sampling summary. The VLS convenience sample and the KP population-based sample were correspondingly healthy older adults. Together, four 10-year age groups were represented, one of which was present in both samples. Sample education differences will be addressed in the analyses.

Measures

Three sets of episodic memory tasks were used: (1) similar categorizable word lists appearing in both the VLS and the KP, (2) a story recall task unique to the VLS, and (3) several tasks unique to the KP (i.e., random word lists, cued recall, and recognition).

Categorizable word list recall: VLS. Six categorizable lists of common English nouns were developed from the Howard (1980) and Battig and Montague (1969) norms. The lists contained 6 words from each of five taxonomic categories, for a total of 30 words. Categories and exemplars were chosen to minimize potential interference effects within and between lists. In general, high-frequency exemplars ranked 2 through 9 were chosen, but to minimize guessing, the most frequently ranked noun was not used. Each participant studied and recalled two lists at each time of measurement. The presentation of lists was counterbalanced across time of measurement to minimize retest effects over time. The lists were presented in unblocked order in typed booklets for study. At presentation, the participants were instructed to remember as many words as possible for later recall, but they were not informed that the words were exemplars of categories. The participants had 2 min to study the words, fol-

lowed immediately by a 5-min period in which they wrote their recalled words on lined paper. The participants were instructed to write down as many of the words as they could remember in any order.

Three principal indicators of performance were used. First, total number of words recalled was scored. In order to be comparable to the word list recall measure of the KP, the total recall score for each of the two lists in each occasion of measurement were averaged and converted to percentage of recall. Second, interference was indexed by the number of intrusions (words from the category that were not on the original list) produced in the recall protocols. Third, the ARC score (Roenker et al., 1971) was computed. This measure provides an index of the degree to which the individual has clustered his or her recall and is typically interpreted in terms of the degree of categorical organization that has been used in encoding and retrieving the words.

Categorizable word list recall: KP. Five lists were randomly distributed to approximately 20% of the participants in each age group. No participant received the same list at both waves. Each list included 12 words, including 3 words from each of four taxonomic categories. The lists were equal with respect to word length, visual and tactual imagery, meaningfulness, and frequency, as determined by a Swedish normative study (Molander, 1984). The words were also typical of their respective categories, according to Swedish norms (Nilsson, 1973).

At presentation, the participants were instructed to remember as many words as possible for purposes of later recall, but they were not informed that the words were exemplars of categories. The words were presented serially at a rate of 5 sec each. They were read aloud by the experimenter and were simultaneously presented printed on cards in 3-cm capital letters. Following presentation of the entire list, the participants were asked to free recall the words, and the experimenter recorded the responses. Two minutes were allowed for free recall. This was followed by a cued recall test in which the category names were presented as cues in a booklet, and the participants were again asked to recall as many words as possible from the study list. Two minutes were allowed for cued recall. As with the VLS data, three principal performance measures (percent total recall, number of intrusions, and ARC score) were evaluated.

Categorizable word lists supplemental measures. Two supplemental indicators of word recall performance were calculated: the percentage of categories recalled (i.e., if at least one word from the category was remembered) and the average percentage of words per category (Tulving & Pearlstone, 1966). Whereas the former has been tentatively associated with retrieval operations, the latter has provisionally been associated with encoding operations. Because previous discussions (Herlitz, Nilsson, & Bäckman, 1997; Small et al., 1999) have explored the extent to which they are linked to gender effects in episodic memory aging, we include the results for follow-up interpretation purposes only in the Discussion section.

Story recall: VLS. Four selections from a set of 25 parallel narratives (Dixon, Hertzog, Friesen, & Hultsch, 1993) were used. They were well organized and structurally equivalent, consisting of 24 sentences, approximately 300 words, and about 160 propositions (idea units). The propositions differed in hierarchical level, depending on their relationship to the themes of the story. The propositions ranged from main ideas (Level 1 propositions) to specific details (Level 7 propositions). For present purposes, these levels were collapsed into four categories: Levels 1 and 2, Level 3, Level 4, and Levels 5–7.

The participants studied and recalled two stories at each time of measurement. The stories were presented in typed booklets for a 4-min study period. The participants were given 10 min to write what they recalled on lined paper. They were instructed to recall as much of the substance of the story as possible, including the main ideas and details. The participants were told that they could recall the story in their own words, those of the story, or any combination thereof. Recall was scored by comparing the participants' protocols with the template text base (list of propositions) representing the

original story (Dixon et al., 1993; Turner & Greene, 1978). Consistent with the instructions, correct generalizations and overspecifications were allowed. The propositions were scored independently of their position in the hierarchy. Scorers were trained by an experienced judge. Seven raters participated in scoring the first-occasion data, whereas six raters participated in scoring the second-occasion data. Reliability estimates on this scoring system, within and across the two occasions, showed that the proportion of agreement was invariably above .90, a value typical for these procedures (e.g., Small et al., 1999).

The number of propositions recalled for each story was converted to proportions. The proportions for each story at each occasion were averaged for total scores. In addition, the proportion recalled at each of the four propositional levels was evaluated. This variable represents the hierarchical structure of the recalled stories.

Additional word list recall tasks: KP. First, two random word lists were prepared, each involving 12 concrete nouns from 12 different semantic categories. The words were taken from a pool of 48 words that were equivalent, as determined from Swedish norms (Molander, 1984). During study, one of the word lists was presented at a rate of 2 sec per word, and the other word list was presented at a rate of 5 sec per word. In the first wave of assessment, the two lists were presented in two orders, each to about half of the participants. At the second wave, the order was reversed for each participant. Preliminary analyses indicated that there was no main effect for order and no interactions involving order. Both tasks were administered in five different versions, according to the same scheme employed for the categorizable lists. The words were read aloud by the experimenter and simultaneously were shown on cards. Oral responses were recorded by the experimenter. Two minutes were allowed for the free recall of each list. In combination with the organizable list recall and the cued recall tasks (described above), these 2- and 5-sec random word tasks may be viewed as occupying four positions on a continuum representing level of cognitive support provided at encoding and retrieval (e.g., Bäckman & Wahlin, 1995). Specifically, study time is added from the 2- to the 5-sec random list; inherent organizational structure is added from the 5-sec random list to free recall of the organizable list; and finally, retrieval support is added from free recall of the organizable words to cued recall of the organizable words.

Second, indicators of primary memory (PM) and secondary memory (SM) were derived from the 2-sec word list data, according to the lag method devised by Tulving and Colotla (1970). This method considers the order in which the items are presented and recalled. An item is assumed to be retrieved from PM if not more than seven items intervene between its presentation and recall. Other items are classified as retrieved from SM.

Third, random word list recognition was also examined. After the free recall of each random word list, yes–no recognition tests were given. The 12 target items were presented again intermixed with an equal number of distractors collected from the same pool of items as the target words. The 24 items were presented consecutively in the same format as during study, and the participants responded orally. Each recognition test took about 2 min to complete. The participants were not informed in advance of the recognition tests.

RESULTS

The results are presented in three segments corresponding to (1) categorizable word recall for both the VLS and the KP samples, (2) VLS story recall performance, and (3) additional KP episodic tasks. Because of the significant age group difference in education in the VLS sample and the apparent differences in educational experience between the VLS and the KP samples, years of education was incorporated as a covariate for all analyses

Table 1
Two-Occasion Performance on Principal Episodic Memory Tasks

Indicators	Study	Age Group	Time 1		Time 2		
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Common Tasks							
% Organized free recall	VLS	YO	63.27	12.64	64.38	12.66	
		MO	62.18	13.13	62.30	14.60	
		OO	51.34	13.39	49.72	14.87	
	KP	OO	57.14	17.84	53.81	19.75	
		VO	52.12	16.19	50.26	18.87	
Intrusions	VLS	YO	0.80	1.17	1.12	1.52	
		MO	0.71	1.12	0.85	1.41	
		OO	0.50	1.05	0.66	1.14	
	KP	OO	0.34	.67	0.33	0.85	
		VO	0.54	0.78	0.29	0.55	
Clustering	VLS	YO	.72	.24	.74	.24	
		MO	.75	.22	.74	.23	
		OO	.66	.27	.68	.27	
	KP	OO	.26	.59	.25	.68	
		VO	.20	.54	.36	.54	
		Other VLS Tasks					
		Proportion of story recall	YO	.39	.10	.38	.09
MO	.38		.09	.37	.09		
OO	.32		.09	.29	.09		
Proportion of levels recall Levels 1 and 2	YO	.48	.12	.55	.16		
	MO	.48	.12	.53	.14		
	OO	.41	.11	.42	.15		
Level 3	YO	.39	.10	.38	.09		
	MO	.38	.10	.36	.09		
	OO	.33	.09	.28	.09		
Level 4	YO	.35	.10	.33	.09		
	MO	.33	.10	.32	.10		
	OO	.26	.09	.25	.10		
Levels 5–7	YO	.30	.13	.29	.12		
	MO	.29	.13	.28	.13		
	OO	.24	.12	.20	.09		
Other KP Tasks							
% Random free recall (2 sec)	OO	44.84	15.78	43.33	16.11		
	VO	41.40	14.03	39.02	14.42		
% Random free recall (5 sec)	OO	49.52	4.97	46.59	15.31		
	VO	43.91	13.97	40.48	17.44		
% Cued recall	OO	73.25	17.25	69.26	18.86		
	VO	68.78	18.20	63.76	17.65		
Recognition (2 sec)	OO	2.91	1.03	2.83	0.94		
	VO	2.76	0.85	2.55	0.96		
Recognition (5 sec)	OO	3.13	0.97	3.19	0.99		
	VO	3.06	0.89	3.09	0.88		
Primary memory	OO	1.95	1.09	1.95	0.94		
	VO	2.17	0.99	2.08	1.07		
Secondary memory	OO	3.38	1.99	3.25	1.92		
	VO	2.63	1.83	2.60	1.62		

Note—VLS, Victoria Longitudinal Study; KP, Kungsholmen Project; YO, young-old; MO, mid-old; OO, old-old; VO, very old. Significant effects are described in the text.

of variance (ANOVAs). The means for all longitudinal gradients and cross-sectional age differences are presented in Table 1.

Categorizable Word Recall

VLS results. A 3 (age group) \times 2 (gender) \times 2 (time) repeated measures analysis of covariance (ANCOVA) was computed on the total percentage of words recalled,

covarying for education level. No significant effect for time of testing was found, thus revealing no overall decline in word recall. Effects for age group [$F(2,393) = 23.55, p < .001; \eta^2 = .11$] and gender [$F(1,393) = 20.93, p < .001; \eta^2 = .05$] were significant. Follow-up tests revealed that the old-old group recalled a lower percentage of words than did the two younger groups, which did not differ from one another ($M_{00} = 50.48, M_{M0} = 60.95,$

$M_{Y0} = 62.89$). Women recalled more words than did men ($M_W = 61.23$, $M_M = 54.98$). The age \times time interaction was not significant ($p > .10$).

Four additional recall measures (number of intrusions and ARC score, as well as the supplemental measures of percentage of categories recalled and percentage of words per category) were examined in a 3 (age group) \times 2 (gender) \times 2 (time) repeated measures multivariate analysis of covariance (MANCOVA). Main effects for age group [Wilks's $\lambda = .874$; $F(8,778) = 6.75$, $p < .001$; $\eta_2 = .07$] and gender [Wilks's $\lambda = .935$; $F(4,389) = 6.76$, $p < .001$; $\eta_2 = .07$] were detected. Notably, the time effect was not significant, and no significant interactions were noted ($ps > .10$). At the univariate level, no effects involving the principal variables of number of intrusions (few such errors were produced) or ARC score (recall protocols were uniformly well organized) were observed ($ps > .10$). The category-related recall results reflected the significant age and gender effects for overall list recall.

KP results. As with the VLS sample, a 2 (age group) \times 2 (gender) \times 2 (time) repeated measures ANCOVA was computed on the percentage of words recalled, covarying years of education. Although the participants recalled a greater percentage of words at the first occasion ($M_{T1} = 54.10$) than at the second ($M_{T2} = 49.42$), this difference, along with the gender ($M_W = 54.39$, $M_M = 49.13$) and age group ($M_{O0} = 49.35$, $M_{VO} = 54.17$) comparisons, were not statistically significant ($ps > .10$). No interactions were significant.

A 2 (age group) \times 2 (gender) \times 2 (time) MANCOVA examined the number of intrusions and the ARC score, as well as the supplemental category-related measures. No significant effects were observed ($ps > .10$).

In sum, an inspection of the longitudinal means in Table 1 reveals that over a 3-year period, very similar patterns of nonsignificant aging effects were observed across the common indicators for both the VLS and the KP. We turn now to the sample-specific episodic tasks.

VLS Story Recall

Two models were examined for text recall. First, a 3 (age group) \times 2 (gender) \times 2 (time) repeated measures ANCOVA, covarying for education, was computed on the total proportion of propositions recalled. As with the word recall data, change in overall recall was modest and not statistically significant. Group differences for age group [$F(2,391) = 18.96$, $p < .001$; $\eta_2 = .09$] and gender [$F(1,391) = 11.24$, $p < .001$; $\eta_2 = .03$] were found. Old-old participants performed at a lower overall level than did the middle-old and young-old participants ($M_{O0} = .30$, $M_{MO} = .37$, $M_{VO} = .38$), and women did better than men ($M_W = .37$, $M_M = .33$). No significant interactions were observed.

Second, a 3 (age group) \times 2 (gender) \times 4 (levels) \times 2 (time) repeated measures ANCOVA was performed to examine group differences in patterns of recall across hierarchical levels, as well as change within levels across time. The significant main effects for age group and gender were consistent with the previous analysis. The main

effect for level of proposition was also significant [$F(3,389) = 20.74$, $p < .001$; $\eta_2 = .14$]. As was expected, recall was highest for the main ideas of the stories, whereas the participants recalled proportionally fewer lower level propositions. The significant age group \times time interaction [$F(2,391) = 4.01$, $p < .02$; $\eta_2 = .02$] was qualified by a marginal age group \times time \times levels interaction [$F(6,778) = 2.09$, $p < .052$; $\eta_2 = .02$]. All three age groups produced relatively well-structured recall protocols at both occasions of measurement. All the groups remembered a somewhat greater number of main ideas at Time 2 than at Time 1. However, whereas the young-old and mid-old groups remembered almost identical proportions of subordinate ideas at both occasions, the old-old adults' recall of details dropped precipitously at Time 2.

Additional KP Measures

Manipulations of cognitive support. The first analysis of the additional KP episodic memory tasks focused on the manipulation of cognitive support. The support factor had four levels, ranging from least support (2-sec random word recall), to somewhat more support (5-sec random word recall), greater support (5-sec organizable word recall), and most support (cued recall). A 2 (age group) \times 2 (gender) \times 2 (time) \times 4 (support) ANCOVA, with repeated measures on the last two factors and covarying for education, was conducted. The main effects for time [$F(1,163) = 5.63$, $p < .001$; $\eta_2 = .03$] and support [$F(3,161) = 21.73$, $p < .001$; $\eta_2 = .29$] and the time \times support interaction [$F(3,161) = 3.16$, $p < .03$; $\eta_2 = .06$] were significant. Both linear [$F(1,163) = 44.04$, $p < .001$] and quadratic [$F(3,163) = 8.90$, $p < .001$] trends were found to be reliable for the support effects. Marginal means for the levels of support at Time 1 ($M_{2sec} = 43.08$, $M_{5sec} = 46.47$, $M_{Org} = 54.09$, $M_{Cued} = 70.40$) and at Time 2 ($M_{2sec} = 39.13$, $M_{5sec} = 43.45$, $M_{Org} = 49.43$, $M_{Cued} = 64.42$) reflect the expected increased recall with more cognitive support and the fact that performance declined reliably over the 3-year period across levels of support. The interaction was due to greater decline for the more than for the less supported conditions, although performance on the former was still superior to performance on the latter at Time 2 (see Table 1).

Recognition tasks. To assess recognition performance, we used d' as an overall measure to control for guessing (Murdock, 1974). A 2 (age group) \times 2 (gender) \times 2 (time) \times 2 (presentation rate) repeated measures ANCOVA was performed, covarying for education. A significant main effect for time was noted [$F(1,163) = 4.12$, $p < .006$; $\eta_2 = .03$]. Marginal means for time ($M_{T1} = 2.99$, $M_{T2} = 2.91$) revealed a lower d' at Time 2 than at Time 1. No significant differences for any other main effects or interactions were observed ($ps > .10$; see Table 1).

Primary and secondary memory. Separate models examining PM and SM over a 3-year period were also tested. Each model was a 2 (age group) \times 2 (gender) \times 2 (time) repeated measures ANCOVA, covarying for education. No significant effects were found for PM ($ps >$

.10). For SM, the time effect was significant [$F(1,163) = 5.70, p < .02; \eta_2 = .03$]. Marginal means indicated that SM declined from Time 1 ($M_{T1} = 3.05$) to Time 2 ($M_{T2} = 2.77$).

DISCUSSION

We have presented longitudinal data from a battery of episodic memory tasks for two independent samples of older adults, ranging in initial age from 54 to 94 years. Our principal goal was to explore generalizability of 3-year memory change patterns across age, sample, and performance indicators. We will address the generalizability issue prior to discussing substantive aspects of episodic memory aging.

Overall, our evidence supports the view that for normal older adults, average memory decline is modest and gradual. This pattern, observed here in two longitudinal studies, appears to apply even across a 40-year swath of older adulthood. Moreover, the detectable decline is similarly gradual for a relatively advantaged convenience sample and a low-education population-based sample. For example, comparing education levels of the common old-old age groups, the VLS OO group had far more schooling than did the KP OO group (about 14 vs. 9 years, respectively). Although population-based samples are reasonably viewed as more representative than convenience samples, our results suggest that similar memory performance and memory change patterns may be observed in some cases. Promoting such comparability are systematic efforts on the part of the researchers using either sampling technique to address the common methodological challenges of cognitive aging noted earlier. These include careful attention to selecting and documenting the characteristics of the sample (e.g., age, health), the design (e.g., testing interval, sample size) and the measures (e.g., multiple indicators). We are aware of one other recent study that has addressed this issue (see Hultsch, MacDonald, Hunter, Maitland, & Dixon, 2002). Although cross-sectional in design, these authors reported few significant differences when contrasting demographic and psychological performance of older adults from a random sample versus two samples of convenience. Specifically, Hultsch et al. (2002) observed no differences on word list recall between the random sample and the convenience sample recruited for a study of memory and aging. Further research will examine the extent to which generalizability across samples applies to longer term normal memory aging or to transitional phases of impaired cognitive aging.

Regarding generalizability across performance indices, the pattern of modest change in performance was clear for the parallel tasks (organized list recall). Indeed, for total words recalled, a classic indicator of episodic memory performance, only one 3-year change was greater than 2%. Moreover, the number of intrusion errors produced was consistently low (all group means were 1.0 or less) for both the VLS and the KP. Although in the VLS the amount increased significantly, given that two 30-

word lists were presented per occasion, such infrequent intrusions may be taken more as evidence of sustained accurate performance than of notable memory distortion. Of course, this is not to say that the VLS adults are immune to memory distortion, for more difficult conditions, such as overlapping categories or faster presentation rates, were not tested. One qualification pertaining to the KP results is that, because the lists were short, the opportunity for errors was probably limited. Nevertheless, we present converging evidence that older adults performed categorizable list recall at a level of high accuracy, in terms of both percentage correct (high) and number of errors (low), and that, on average, these levels are sustained over 3 years, even in very old age.

One possible methodological concern derives from the fact that we examined 3-year change by using 10-year age bands for our cohort groups. Would the patterns and interpretations be different if 3-year change was evaluated for groups with 3-year age bands? In order to compare informally cross-sectional differences with longitudinal changes, we reconstituted the data set to reflect multiple 3-year groups. For the VLS, there were 10 such groups (covering the band from 54 to 83 years), and for the KP, there were 6 such groups (covering the band from 75 to 92). An ANOVA on the VLS data revealed a reliable age effect differentiating the 7 pre-75-year-old groups from the 3 post-75-year-old groups. The 6 post-75-year-old KP groups did not differ from one another. As was expected from previous analyses, no time (decline) effects were observed for either sample. In addition, the mid-70s also differentiated the groups' longitudinal change patterns. The VLS pre-75 groups had an average change of 0.77%, indicating overall stability. In contrast, for the post-75 groups, the change values were $M = -2.76\%$ (VLS) and $M = -2.20\%$ (KP), both indicating losses. Interestingly, the two oldest KP groups showed virtually no change, suggesting, perhaps, that those who have survived relatively intact to the 75- to 85-year period may not be subjected to continued decline.

A further follow-up question may be posed: Would these 3-year change patterns have been predicted from 3-year cross-sectional comparisons? Calculating Time 1 differences between contiguous groups for the VLS pre-75-year-olds resulted in an overall $M = -.40\%$, suggesting a slight loss. For the post-75 comparisons, larger average group differences of $M = -2.54\%$ (VLS) and $M = -1.32\%$ (KP) were obtained. These supplemental comparisons suggest three issues for future research: (1) Differential change patterns may occur around the mid-70s, (2) a leveling of loss may occur after this transition phase, and (3) within this age band, 3-year age comparisons produce similar but imprecise estimates of actual 3-year change.

The ARC measure indicates the extent to which individuals cluster the words they remember by semantic category. Spontaneously identifying and using the organizational characteristics of lists may lead to overall recall benefits, inasmuch as remembering an exemplar of

a given category may lead to further directed effort to recall category items. Clustering was quite high for all the groups and showed no tendency to decline for either sample. The VLS ARC estimates were somewhat higher than those for the KP but were similar in value to those obtained with another sample from a previous longitudinal sequence (Small et al., 1999). Similarly, the KP values correspond to those obtained with similar lists and older Swedish adults (Bäckman & Larsson, 1992). The apparent difference between the VLS and the KP samples may partly reflect the greater number of items present in the VLS lists, a factor that influences ARC estimates (Roemer et al., 1971), as well as the possibility that the written recall format (as required in the VLS) may offer greater opportunity for clustering than does the KP's oral recall format.

The VLS's story recall task provided unique recall materials that were more verbally elaborated than word lists. Although again no reliable 3-year time effects were observed, the present oldest adults (i.e., the old-old group) recalled less information than did the two younger groups. As was noted with word recall, age (or cohort) of the participant, even within samples of positively selected older adults, matters in episodic memory performance. Researchers using one group of older adults as a single point of comparison with younger adults or with special populations of older adults are cautioned that the mean age and age range of the sample are important concerns (see Bäckman et al., 2000; Hertzog, 2004).

The prose passages contained propositions differing in level of thematic content. Previous research in which relative proportions of information recalled from thematic levels have been examined has revealed that the recalled information is relatively well structured for both younger and older adults (e.g., Dixon & Gould, 1998; Stine, Soederberg, & Morrow, 1996). The present young-old and mid-old adults produced nearly identical profiles of recall across levels of information at both occasions. By comparison, the old-old adults' profiles were at a consistently lower level of performance. Moreover, the marginally significant ($p < .052$) three-way interaction (age \times time \times levels) suggests the potential emergence of an important shift in late life. At Time 2, the old-old adults maintained a relatively high proportion of main idea recall, but marginally diminished levels of detail recall. Several observers have noted that older adults may actively focus narrative recall attention on superordinate ideas as a means of compensating for decrements in overall recall (e.g., Gould, Trevithick, & Dixon, 1991). Not yet known, but deserving further investigation, is whether such adaptive mechanisms originate in deliberate or strategic efforts or whether the phenomenon is linked to aging-sensitive differential recall-enhancing features of structural elements in prose passages. If main ideas may serve as retrieval structures (promoting recall of linked subordinate ideas; e.g., van Dijk & Kintsch, 1983), then even quite old adults maintain the ability to

identify and produce them over a period of 3 years. With further waves of measurement in the VLS, it will be interesting to test whether, during these decades, recall of details may be especially sensitive to the effects of aging (see also Small et al., 1999).

The additional episodic measures available in the KP provide a unique opportunity to examine whether the effects of cognitive support on memory performance are (1) present in very old adults and (2) maintained across a 3-year interval. As can be seen in Table 1, more support (presence of cues) was associated with better recall performance than was less support for both age groups. Interestingly, significant decline was observed, but these two main effects were modified by an interaction. An examination of the means suggests that there is a tendency for the decline to be greater for cued recall (the most supported task) than for the other three levels of cognitive support. Whereas the decline for cued recall performance was approximately 6%, the decline for the other three levels was between 3% and 4.5%. To be sure, the performance levels at both occasions are clearly greater for the cued recall task (between 70% at Time 1 and 64% at Time 2) than for the other tasks (ranging from 39% to 54%). The results of the recognition task, which is also considered as cognitively supported, confirm this observation.

The potential for benefitting from the presence of cognitive support may be less available in very late life than it is in early and middle adulthood (Bäckman et al., 2000). Put another way, more supportive task conditions (as compared with the less supportive conditions) may offer a more sensitive test of episodic memory change in old age. One possible theoretical reason is that very old adults may have already experienced the normal decline that occurs in relatively unsupported episodic tasks. This line of reasoning is consistent with the view that there are aging-related reductions in cognitive reserve capacity, or that the potential of cognitive performance may gradually decline through late life (Baltes, Staudinger, & Lindenberger, 1999). In this sense, being very old represents one of several conditions systematically associated with memory deficits and also with reductions in the size of performance gains from the provision of cognitive support. Another possible methodological consideration is that this differential decline in supported episodic performance may be especially notable when contrasted with the least supported tasks used here. The present least supported exemplars combine a speeded manipulation (2- and 5-sec presentation rates) with random word lists. Tasks demanding speeded performance in old age are known to produce robust aging effects (e.g., Salt-house, 1991). Thus, the two speeded tasks are difficult for very old adults not only because of lack of cognitive support, but also because they have already experienced accumulating losses in speed of processing. The randomness of the lists contributes to the complexity, although the relative brevity (12 words) makes it possible to remember about 40% (i.e., 4 or 5 words) of the list.

Very old adults, therefore, may be operating at basic or functional (working memory?) capacity in such tasks and, thus, have little available room to decline further. Whereas basic unsupported and speeded tasks are sensitive to aging-related changes earlier in life, more supported tasks may be where most of the action is in very late life.

Although the short-term decline in performance on the more supported tasks is theoretically compelling, it should be noted that some decline occurred also for the least supported tasks. Specifically, whereas no decline was observed for the PM component in recall of the two random word lists, small but significant decline was found for the SM component of these tasks. The finding that PM, which indexes a relatively passive holding of information in consciousness, is well preserved in very old age is consistent with previous research (e.g., Wahlin, Bäckman, & Winblad, 1995). That SM deteriorated slightly during the 3-year follow-up period suggests that the ability to transfer information from consciousness to some form of permanent memory representation undergoes gradual late-life decline (cf. Craik, 1983). Arguably, most normal age-related decline in performance on tasks that involve paced encoding conditions and unstructured memory lists may occur earlier in life. To this observation, we add that additional decline may be expected for performance on tasks in which the ability to establish new episodic representations is tapped.

Although the focus of this research has been on time and age effects, the role of gender in the aging of episodic memory performance has received recent attention. Indeed, gender effects in older adults have been observed on numerous measures of brain structure, brain functioning, cognition, and even everyday competence (e.g., Meinz & Salthouse, 1998; Smith & Baltes, 1998). Although at the neurological level gender comparisons may yield interactions indicating more rapid decline in men than in women (e.g., Coffey et al., 1998), at the cognitive level selected main effects are more typically observed (e.g., Larrabee & Crook, 1993; Schaie, 1996; Small et al., 1999). In a large cross-sectional sample of adults stratified by decade, Herlitz et al. (1997) reported gender differences for episodic memory performance, but not for other aspects of memory (e.g., semantic memory). Like Herlitz et al., we observed several instances in the VLS data (but not in the KP data) in which older women recalled more information than did older men. However, despite the wide age range of older adults and the opportunity for observing longitudinal change, little evidence for decline (i.e., time effects) and certainly not accelerated decline for older men (i.e., gender \times time interactions) was detected. Herlitz et al. speculated that episodic gender differences may be due more to encoding than to retrieval deficits in men. Given our observed gender differences in VLS word recall, we explored this hypothesis by using the two supplemental measures available for categorizable word lists. Specifically, we tested gender differences in the ancillary performance measures of percentage of

words per category (encoding-related) and percentage of categories recalled (retrieval-related) variables. According to these measures, gender differences were not linked more to encoding than to retrieval operations; men performed significantly more poorly than women on both of these variables. In sum, gender effects in episodic memory are small but often significant in large samples (e.g., Herlitz et al., 1997; Hultsch et al., 1998; but see Zelinski & Burnight, 1997). The present data did not support the idea that very old men may be showing accelerated episodic memory decline (cf. Meinz & Salthouse, 1998). Detecting this phenomenon may require longer term change data.

Two limiting factors of the present research should be noted. A first consideration, noted above, is the 3-year interval between occasions of measurement. Naturally, more substantial cognitive changes in adults may be expected over longer periods. Nevertheless, the VLS and KP designs were informed by the principle that 3 years may be an ideal interval for detecting and charting minor changes, especially in late life. Whereas 3 years is a short period in the context of the life span or even adulthood, it is not so short in the context of late adulthood, when the expected years-to-live is measured in single digits. Furthermore, substantial changes occur in many fundamental aspects of cognition (such as verbal processing speed and working memory) over such intervals in late life (e.g., Hultsch et al., 1998). In addition, we detected interpretable decline in some aspects of episodic memory. Thus, in the continuing discussion of how best to chart cognitive changes in adulthood, we add to the well-known issue of interval length the relatively new factors of task characteristics (e.g., cognitive support) and late-life age considerations (i.e., occasional differential rates of change).

A second limiting factor applies to all studies of cognitive change; specifically, practice and selective survival may tend to enhance the level of observed performance for longitudinal survivors (e.g., Perls, Morris, Ooi, & Lipsitz, 1993). As was noted, both the VLS and the KP follow typical methodological recommendations to minimize practice effects (e.g., multiple sets of equivalent materials and relatively long intervals). Selective attrition, however, is present in virtually all such designs. In contrast, terminal decline and preclinical dementia processes, which are possibly operating even in seemingly normal very old adults, would have the effect of lowering individual performances (e.g., Bäckman et al., 2002; Berg, 1996; Sliwinski et al., 1996). Although the present samples were deliberately selected so as to eliminate any known preclinical condition, only future retestings can confirm this selection. The aim of studying normal cognitive aging is complicated by numerous factors, and no uncompromising or perfect method exists (Hertzog & Dixon, 1996). Among the more promising methods, however, are those that produce data pertaining to actual cognitive change.

REFERENCES

- BÄCKMAN, L., JONSSON LAUKKA, E., WAHLIN, Å., SMALL, B. J., & FRATIGLIONI, L. (2002). Influences of preclinical dementia and impending death on the magnitude of age-related cognitive deficits. *Psychology & Aging*, *17*, 435-442.
- BÄCKMAN, L., & LARSSON, M. (1992). Recall of organizable words and objects in adulthood: Influences of instructions, retention interval, and retrieval cues. *Journal of Gerontology: Psychological Sciences*, *47*, 273-278.
- BÄCKMAN, L., SMALL, B. J., WAHLIN, Å., & LARSSON, M. (2000). Cognitive functioning in very old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of cognitive aging* (2nd ed., pp. 499-558). Mahwah, NJ: Erlbaum.
- BÄCKMAN, L., & WAHLIN, Å. (1995). Influences of item organizability and semantic retrieval cues on word recall in very old age. *Aging & Cognition*, *2*, 312-325.
- BALTES, P. B., STAUDINGER, U. M., & LINDENBERGER, U. (1999). Life span psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*, 471-507.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category names for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*(3, Pt. 2), 1-46.
- BERG, S. (1996). Aging, behavior, and terminal decline. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 323-337). New York: Academic Press.
- COFFEY, C. E., LUCKE, J. F., SAXTON, J. A., RATCLIFF, G., UNITAS, L. J., BILLIG, B., & BRYAN, R. N. (1998). Sex differences in brain aging. *Archives of Neurology*, *55*, 169-179.
- COLSHER, P. L., & WALLACE, R. B. (1991). Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, *1*, 215-230.
- 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.
- DIXON, R. A., & GOULD, O. N. (1998). Younger and older adults collaborating on retelling everyday stories. *Applied Developmental Science*, *2*, 160-171.
- DIXON, R. A., HERTZOG, C., FRIESEN, I. C., & HULTSCH, D. F. (1993). Assessment of intraindividual change in text recall of elderly adults. In H. H. Brownwell & Y. Joanne (Eds.), *Narrative discourse in neurologically impaired and normal aging adults* (pp. 77-101). San Diego: Singular.
- FRATIGLIONI, L., VIITANEN, M., BÄCKMAN, L., SANDMAN, P.-O., & WINBLAD, B. (1992). Occurrence of dementia in advanced age: The study design of the Kungsholmen project. *Neuroepidemiology*, *11* (Suppl. 1), 29-36.
- GIAMBRA, L. M., ARENBERG, D., ZONDERMAN, A. B., KAWAS, C., & COSTA, P. T. JR. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology & Aging*, *10*, 123-139.
- GOULD, O. N., TREVITHICK, L., & DIXON, R. A. (1991). Adult age differences in elaborations produced during prose recall. *Psychology & Aging*, *6*, 93-99.
- HERLITZ, Å., NILSSON, L.-G., & BÄCKMAN, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, *25*, 801-811.
- HERTZOG, C. (2004). Does longitudinal evidence confirm theories of cognitive aging as derived from cross-sectional data? In R. A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 41-64). Oxford: Oxford University Press.
- HERTZOG, C., & DIXON, R. A. (1996). Methodological issues in research on cognition and aging. In F. Blanchard-Fields & T. M. Hess (Eds.), *Perspectives on cognitive change in adulthood and aging* (pp. 66-121). New York: McGraw-Hill.
- HOWARD, D. V. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology*, *35*, 884-890.
- HULTSCH, D. F., HERTZOG, C., DIXON, R. A., & SMALL, B. J. (1998). *Memory change in the aged*. New York: Cambridge University Press.
- HULTSCH, D. F., MACDONALD, S. W. S., HUNTER, M. A., MAITLAND, S. B., & DIXON, R. A. (2002). Sampling and generalizability in developmental research: Comparison of random and convenience samples of older adults. *International Journal of Behavioral Development*, *26*, 345-359.
- LARRABEE, G. J., & CROOK, T. H. (1993). Do men show more rapid age-associated decline in simulated everyday verbal memory than do women? *Psychology & Aging*, *8*, 68-71.
- MACDONALD, S. W. S., DIXON, R. A., COHEN, A.-L., & HAZLITT, J. E. (2004). Biological age and 12-year cognitive change in older adults: Findings from the Victoria Longitudinal Study. *Gerontology*, *50*, 64-81.
- MCCARTY, S. M., SIEGLER, I. C., & LOGUE, P. E. (1982). Cross-sectional and longitudinal patterns of three Wechsler Memory Scale subtests. *Journal of Gerontology*, *37*, 169-175.
- MEINZ, E. J., & SALTHOUSE, T. A. (1998). Is age kinder to females than to males? *Psychonomic Bulletin & Review*, *5*, 56-70.
- MOLANDER, B. (1984). *Imagery, visual and tactual dimensions of imagery and meaningfulness: Swedish norms for 858 nouns* (Umeå Psychological Rep. No. 178). Umeå, Sweden: University of Umeå, Department of Psychology.
- MURDOCK, B. B., JR. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.
- NILSSON, L.-G. (1973). *Category norms for verbal material* (Uppsala Psychological Rep. No. 135). Uppsala, Sweden: University of Uppsala, Department of Psychology.
- PERLS, T. T., MORRIS, J. N., OOI, W. L., & LIPSITZ, L. A. (1993). The relationship between age, gender and cognitive performance in the very old: The effect of selective survival. *Journal of the American Geriatrics Society*, *41*, 1193-1201.
- ROENKER, D. L., THOMPSON, C. P., & BROWN, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45-48.
- SALTHOUSE, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.
- SCHAE, K. W. (1996). *Intellectual development in adulthood*. New York: Cambridge University Press.
- SCHAE, K. W., & HOFER, S. M. (2001). Longitudinal studies of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 53-77). San Diego: Academic Press.
- SLIWINSKI, M., LIPTON, R. B., BUSCHKE, H., & STEWART, W. (1996). The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *Journal of Gerontology: Psychological Sciences*, *51*, 217-225.
- SMALL, B. J., DIXON, R. A., HULTSCH, D. F., & HERTZOG, C. (1999). Longitudinal changes in quantitative and qualitative indicators of word and story recall in young-old and old-old adults. *Journal of Gerontology: Psychological Sciences*, *54B*, P107-P115.
- SMITH, J., & BALTES, M. M. (1998). The role of gender in very old age: Profiles of functioning and everyday life patterns. *Psychology & Aging*, *13*, 676-695.
- STATISTISKA CENTRALBYRÅN (1997). *Befolkningens utbildningsnivå* [Educational levels of inhabitants]. Stockholm: Author.
- STINE, E. A. L., SOEDERBERG, L., & MORROW, D. G. (1996). Language and discourse processing through adulthood. In F. Blanchard-Fields & T. M. Hess (Eds.), *Perspectives on cognition in adulthood and aging* (pp. 255-290). New York: McGraw-Hill.
- TAYLOR, J. L., MILLER, T. P., & TINKLENBERG, J. R. (1992). Correlates of memory decline: A 4-year longitudinal study of older adults with memory complaints. *Psychology & Aging*, *7*, 185-193.
- TULVING, E., & COLOTLA, V. (1970). Free recall of trilingual lists. *Cognitive Psychology*, *1*, 86-98.
- TULVING, E., & PEARLSTONE, Z. (1966). Availability versus access of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, *5*, 381-391.
- TURNER, A., & GREENE, F. (1978). Construction and use of a propositional text base. *JSAS: Catalog of Selected Documents in Psychology*, *8* (MS. No. 1713).
- VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- WAHLIN, Å. (2004). Health, disease, and cognitive functioning in old

- age. In R. A. Dixon, L. Bäckman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 279-302). Oxford: Oxford University Press.
- WAHLIN, Å., BÄCKMAN, L., & WINBLAD, B. (1995). Free recall and recognition of slowly and rapidly presented words in very old age: A community-based study. *Experimental Aging Research*, **21**, 251-271.
- ZACKS, R. T., HASHER, L., & LI, K. Z. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 293-358). Mahwah, NJ: Erlbaum.
- ZELINSKI, E. M., & BURNIGHT, K. P. (1997). Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. *Psychology & Aging*, **12**, 503-513.
- ZELINSKI, E. M., GILEWSKI, M. J., & SCHAIE, K. W. (1993). Individual differences in cross-sectional and 3-year longitudinal memory performance across the adult life span. *Psychology & Aging*, **8**, 176-186.

(Manuscript received December 28, 1998;
revision accepted for publication June 4, 2004.)