

Selective sex differences in declarative memory

SCOTT B. MAITLAND

University of Guelph, Guelph, Ontario, Canada

AGNETA HERLITZ

Aging Research Center, Karolinska Institute, Stockholm, Sweden

LARS NYBERG

Umeå University, Umeå, Sweden

LARS BÄCKMAN

Aging Research Center, Karolinska Institute, Stockholm, Sweden

and

LARS-GÖRAN NILSSON

University of Stockholm, Stockholm, Sweden

Sex invariance of a six-factor, higher order model of declarative memory (two second-order factors: episodic and semantic memory; and four first-order factors: recall, recognition, fluency, and knowledge) was established for 1,796 participants (35–85 years). Metric invariance of first- and second-order factor loadings across sex was demonstrated. At the second-order level, a female advantage was observed for both episodic and semantic memory. At the first-order level, sex differences in episodic memory were apparent for both recall and recognition, whereas the differences in semantic memory were driven by a female superiority in fluency. Additional tests of sex differences in three age groups (35–50, 55–65, and 70–85 years of age) indicated that the female superiority in declarative memory diminished with advancing age. The factor-specific sex differences are discussed in relation to sex differences in hippocampal function.

In memory everything is preserved separately, according to its category.

—St. Augustine, *Confessions*, 10.8
(tr. R.S. Pine-Coffin, 1961)

Potential sex differences have been examined in a variety of cognitive domains (e.g., Feingold, 1996; Halpern & LaMay, 2000; Hyde, Fennema, & Lamon, 1990; Hyde & Linn, 1988; Maitland, Intrieri, Schaie, & Willis, 2000;

Rosén, 1995; Voyer, Voyer, & Bryden, 1995). For example, recent research has addressed the existence and generality of sex differences in declarative memory. There is emerging evidence for a female superiority in episodic memory (e.g., Herlitz, Airaksinen, & Nordström, 1999; Herlitz, Nilsson, & Bäckman, 1997; Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997), as well as in some aspects of semantic memory (e.g., verbal fluency; Herlitz et al., 1997; Hyde & Linn, 1988). Nevertheless, the extent to which the female advantage generalizes across different aspects of declarative memory remains unclear. A major issue in this context is that systematic comparisons of the magnitude of sex differences for tests of episodic memory, such as recall and recognition, are largely missing. In addition, it remains unclear whether sex differences, whenever observed, are invariant across the adult life span. Therefore, the overall purpose of the present research was to conduct a large-scale investigation of sex differences in declarative memory in adulthood and old age.

The starting point for the current investigation was a six-factor model of declarative memory (Nyberg et al., 2003). This model includes higher order factors for episodic and semantic memory and first-order factors for recall, recognition, knowledge, and fluency, respectively (see Figure 1). This model was selected as fitting the data better than unitary (declarative memory), two-factor

This research was supported by Grant 97/1821 from the Swedish Foundation for International Cooperation in Research and Higher Education to L.-G.N. and Roger Dixon. This grant is designed to promote collaborative research between the Betula Longitudinal Study, funded by the Bank of Sweden Tercentenary Foundation (1988-0082:17), the Swedish Council for Planning and Coordination of Research (D1988-0092, D1989-0115, D1990-0074, D1991-0258, D1992-0143, D1997-0756, D1997-1841, D1999-0739, B1999-474), the Swedish Council for Research in the Humanities and Social Sciences (F377/1988–2000), and the Swedish Council for Social Research (1988–1990: 88-0082, and 311/1991–2000), and the Victoria Longitudinal Study, funded by the National Institute on Aging (AG08235). The present data were taken from the Betula Study. A.H., L.B., and L.N. are supported by the Swedish Council for Research in the Humanities and Social Sciences. We also acknowledge the contribution of the staff of the Betula Project. Correspondence concerning this article should be addressed to S. B. Maitland, Family Relations and Applied Nutrition, University of Guelph, Guelph, ON, Canada, N1G 2W1 (e-mail: smaitlan@uoguelph.ca).

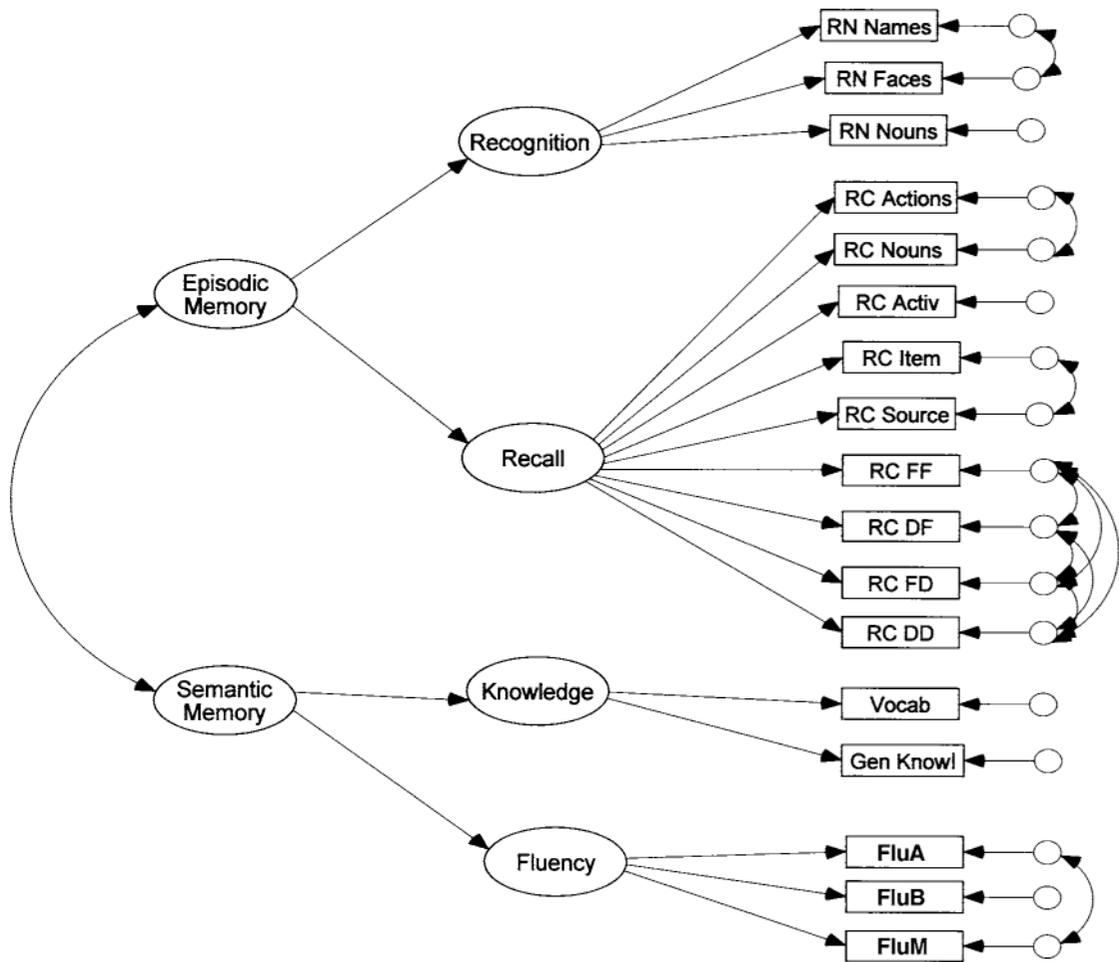


Figure 1. Conceptual model defined by Nyberg et al. (2003).

(episodic, semantic), and four-factor (recall, recognition, knowledge, fluency) solutions. In particular, the six-factor model has been demonstrated to have more parsimony and marginally better fit than the four-factor solution. An additional benefit of the six-factor model is the ability to examine group differences in factor means at two levels, including the subordinate first-order factors and the superordinate episodic and semantic factors. Another important aspect of the model is that the factors are represented by multiple measures for each subtype of memory tested. The validity of the model has been established by confirmatory factor analysis, and the model was found to provide a good account of the structure of declarative memory for differently aged persons across adulthood.

In this study we sought to examine the existence of sex differences at hierarchical levels of this model. More generally, we examined sex differences in episodic and semantic memory. At a more detailed level, we examined differences at the level of first-order factors (i.e., recognition, recall, knowledge, fluency). A significant feature of the present study is that before examining the existence

of sex differences in memory performance, we explored whether the six-factor model provided an adequate representation of the data for both women and men (i.e., whether the model was sex invariant). The importance of determining that underlying measurement structures are age invariant before making quantitative comparisons of younger and older adults' test performance has been emphasized in the cognitive aging literature (e.g., Anstey, Hofer, & Luszcz, 2003; Baltes, Reese, & Nesselroade, 1988; Hertzog & Nesselroade, 2003; Horn & McArdle, 1992; Labouvie, 1980; Nyberg et al., 2003; Schaie, Maitland, Willis, & Intrieri, 1998). Lacking this evidence, one would not be assured of measuring the same construct across age groups. Analogously, quantitative comparisons of cognitive test scores of men and women rest on the assumption that the underlying measurement structure is sex invariant, but very few studies have explicitly tested this assumption (for exceptions, see Maitland et al., 2000; Rosén, 1995).

Therefore, sex invariance of the multifactor model was examined across multiple age groups to determine whether

aging influences the structure as well as the patterns of sex differences in declarative memory. The present study thus had two main objectives: First, we sought to establish evidence for sex invariance in the structure of declarative memory. Second, we determined whether empirical patterns of sex differences in memory performance demonstrated in the literature at the manifest level could be observed at the level of latent constructs. These goals were accomplished using data from a population-based sample of healthy individuals ranging in age between 35 and 85 years.

METHOD

Participants

We examined 1,796 community-dwelling individuals from the Betula Project (BP; Nilsson et al., 1997). Participants were drawn randomly from the population registry in Umeå, a city of about 100,000 inhabitants in northern Sweden. Exclusion criteria were dementia, serious visual and auditory handicaps, and a mother tongue other than Swedish. The BP involves multiple independent samples. Participants from Samples 2 and 3 were combined for this study after comparisons indicated similar demographic and performance profiles. These participants were sampled during the same time period, 1993–1995. The sampling procedure used was identical to that used when participants were sampled for another independent subgroup (Sample 1) 5 years earlier (1988–1990). The sample, aged 35–85 years, was divided by sex, resulting in 1,005 women and 791 men ($M_{\text{age-women}} = 60.03$ years, $M_{\text{age-men}} = 57.47$ years). Three age groups were examined: 671 middle-aged adults between 35 and 50 years (343 women, $M_{\text{age}} = 43.75$; 328 men, $M_{\text{age}} = 43.46$), 569 young-old adults between 55 and 65 years (318 women, $M_{\text{age}} = 60.15$; 251 men, $M_{\text{age}} = 59.77$), and 556 older adults between 70 and 85 years (344 women, $M_{\text{age}} = 75.68$; 212 men,

$M_{\text{age}} = 75.63$). Previous reports have shown that the majority of participants in the BP report themselves as healthy (Nilsson et al., 1997). Furthermore, it has been noted that level of education becomes significantly lower across age groups and that Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) scores below 24 increase with age (Nyberg et al., 2003). Therefore, we covaried by education in all analyses, and all participants with MMSE scores below 24 were excluded in order to control for possible early dementia (e.g., Small, Viitanen, & Bäckman, 1997).

Measurement Variables

The memory tasks used were selected to cover a wide range of memory systems and processes for the sake of evaluating developmental effects across adulthood and old age. The test battery in BP included tasks to assess episodic memory, semantic memory, primary memory, and priming. In order to assess processes at encoding and retrieval separately, the amount of cognitive support was manipulated at these two stages. Admittedly, the majority of the tasks in the study were of a verbal character, but nonverbal tasks were also included—for example, face recognition. The memory tasks were administered during two test sessions, lasting between 1.5 and 2 h each (see Nilsson et al., 1997; also see Nilsson, 1999, for a detailed description). In the following descriptions, letters within parentheses refer to the names of measures entered in the analyses. Descriptive information for all comparison groups is provided in Table 1.

Episodic Tasks: Recall

Word recall under focused and divided attention incorporated recall under four conditions of focused and divided attention at encoding and retrieval. Subjects were presented auditorily with a list of 12 common, unrelated nouns at a 2-sec rate and instructed to learn the words for an immediate free-recall test. Two sets of four lists of words were used. Mean word frequency for the four lists was 98.6 words per million words written text in one set and 98.5 words per million in the other (Allén, 1972). Frequency ranges were very similar for the eight lists, between 50 and 200 words per million.

Table 1
Means and Standard Deviations for All Tests/Groups

Variable	All Women		All Men		35–50 years				55–65 years				70–85 years			
					Women		Men		Women		Men		Women		Men	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Recognition																
RN Names	3.92	2.24	3.73	2.36	4.65	2.16	4.34	2.43	3.98	2.08	3.54	2.23	3.12	2.20	3.00	2.15
RN Faces	6.42	3.03	5.72	3.07	7.89	2.51	7.00	2.76	6.62	2.82	5.61	2.79	4.78	2.88	3.86	2.85
RN Nouns	6.72	1.41	6.56	1.63	7.14	1.15	6.96	1.37	6.87	1.16	6.61	1.35	6.16	1.65	5.87	2.02
Recall																
RC Actions	8.51	3.10	8.29	3.05	10.42	2.36	9.99	2.51	8.76	2.72	8.12	2.44	6.36	2.72	5.86	2.74
RC Nouns	9.99	2.79	9.70	2.76	11.62	2.04	10.87	2.32	10.24	2.40	9.85	2.39	8.14	2.68	7.74	2.73
RC Activ	11.10	4.30	10.72	4.11	14.00	3.87	13.08	3.94	10.91	3.50	9.71	2.98	8.39	3.48	8.26	3.54
RC Item	5.92	3.10	5.27	2.98	7.06	3.18	6.35	2.89	6.12	3.00	5.34	3.01	4.61	2.59	3.52	2.13
RC Source	4.68	2.86	4.42	2.69	6.09	2.65	5.52	2.65	4.98	2.61	4.50	2.41	3.00	2.38	2.60	2.04
RC FF	5.39	1.61	5.14	1.60	6.05	1.57	5.71	1.54	5.74	1.36	5.05	1.48	4.40	1.39	4.38	1.49
RC DF	3.98	1.39	3.90	1.31	4.52	1.31	4.41	1.20	4.15	1.22	3.84	1.21	3.28	1.33	3.18	1.24
RC FD	4.63	1.63	4.35	1.53	5.35	1.54	4.99	1.43	4.86	1.40	4.29	1.31	3.69	1.47	3.43	1.46
RC DD	3.73	1.44	3.57	1.30	4.27	1.42	3.98	1.21	3.83	1.26	3.66	1.24	3.09	1.38	2.81	1.19
Knowledge																
Vocab	22.08	5.08	21.82	4.96	23.97	3.96	23.30	3.80	22.89	4.44	21.73	4.88	19.45	5.53	19.63	5.76
Gen Knowl	7.56	1.80	7.70	1.64	7.99	1.56	7.91	1.60	7.90	1.64	7.87	1.53	6.80	1.92	7.18	1.73
Fluency																
FluA	11.57	4.76	10.71	4.60	13.15	4.47	12.22	4.56	12.03	4.55	10.26	4.30	9.58	4.54	8.90	4.25
FluB	4.65	2.18	4.49	2.19	5.26	2.19	5.00	2.09	5.10	2.08	4.48	2.13	3.62	1.87	3.74	2.21
FluM	5.92	3.10	5.04	3.03	6.60	3.12	5.45	2.98	6.47	3.01	5.10	3.13	4.75	2.83	4.34	2.88
Education																
	10.06	3.92	10.45	4.04	12.98	3.23	12.95	3.42	9.78	3.61	9.26	3.42	7.41	2.61	7.98	3.42

Word length varied between 3 and 7 letters, and mean word length varied from 4.75 to 5.10 letters between lists. The eight lists were counterbalanced across the four conditions and the subjects in each age cohort. The participants were allowed 1 min for oral free recall. Word recall was paired with a concurrent task, sorting a deck of black and red cards into two stacks, one red and one black. Using this sorting task, we created the four conditions of divided versus focused attention: The concurrent task took place during study only (RC DF), during test only (RC FD), during both study and test (RC DD), or during neither study nor test (RC FF). The word recall task was selected as an episodic memory task with or without influence of divided attention, a choice in line with general consensus in the field. However, this task can also be conceived of as a working memory task when the differences between conditions and the performance in the card-sorting task per se are taken into account (Baddeley, Lewis, Eldridge, & Thomson, 1984). We did not approach the task in this way.

For *recall of actions*, participants were presented with two lists of 16 verb–noun sentences, each denoting a simple action (e.g., lift the book). Each sentence was presented on an index card for 8 sec. The nouns of the sentences in each list belonged to four semantic categories (e.g., fruits, musical instruments) with four different exemplars of each category. Mean percentages in category norms for the two lists were 28.3 (range, 0–88) and 28.5 (range, 0–84), respectively. For one list, participants were requested to enact each sentence using the specified object. The other list was studied without enactment. A free-recall test of the sentences followed (RC Actions). The participants were allowed 2 min for oral free recall.

Following the free-recall test, participants were provided with a sheet listing the eight semantic categories into which the nouns of the 32 sentences (described above) could be divided. *Cued recall of nouns* required the participants to remember the nouns, with the category names serving as cues. Number of nouns recalled from the enacted encoding condition (RC Nouns) served as a measure in the analyses.

In *recall of newly acquired facts*, subjects were presented with 20 fictitious statements concerning famous people (e.g., “Astrid Lindgren collects stamps as a hobby.”) presented in four ways: auditorily (in a female or male voice) or visually (printed on a yellow or red card). At test, the subjects were asked questions about the fictitious statements (e.g., “What is Astrid Lindgren’s hobby?”) (RC Item), and they were also asked to identify the medium through which the information was originally presented (RC Source). A version of the source recall task was also presented with the general knowledge task (see below). There, participants were given the option to describe how they had learned a piece of information extra-experimentally; in school, in papers or books, on radio or television, through neighbors or friends, or in some other way.

Activity recall required the subjects at the end of the testing session to incidentally recall all tasks they had performed (RC Activ).

Episodic Tasks: Recognition

Face and name recognition consisted of a one-by-one presentation of 16 color photos of faces of 10-year-old children, together with a name (first and family name). Later, subjects were presented consecutively with 12 target faces and 12 distractor faces in a random order for a *yes/no* recognition test (RN Faces). Presentation rate was 8 sec per picture. Participants were also asked to identify the names from a list of four alternatives (RN Names).

In *recognition of nouns*, participants were presented with 32 nouns, including 16 from the enacted/nonenacted sentences studied earlier and 16 distractors, and were asked to do a *yes/no* recognition judgment (RN Nouns).

Semantic Tasks: Knowledge

For the *vocabulary* measure, a 30-item multiple-choice synonym test (Dureman, 1960) was used. Total number of correctly identified synonyms (Vocab) was entered into the analyses.

The 10 questions included in the *general knowledge* task were possible to answer on the basis of acculturated knowledge (e.g., “Where was Ingemar Stenmark born?”) (Gen Knowl). As mentioned above, participants were given the chance to identify the source of their knowledge after each question they answered.

Semantic Tasks: Fluency

In order to measure their *word fluency*, participants were asked to generate as many words as possible during 1 min for each of three criteria: (1) words with the initial letter *A* (FluA); (2) five-letter words with the initial letter *M* (FluM); and (3) names of professions with an initial letter *B* (FluB). The word fluency task is sometimes also considered a working memory task, especially when initial letter is a criterion for word generation. A fourth fluency test was also given, although it is not included in the analyses reported here due to possible floor effects. In this task, subjects were asked to generate five-letter names of animals with the initial letter *S*.

Statistical Procedures

Models were tested using LISREL 8.51 (Jöreskog & Sörbom, 2001). Analyses were conducted on covariance matrices and mean vectors, with results of the final models reported as standardized estimates for ease of interpretation. Factor scaling was accomplished by fixing one item for each factor (either first- or second-order) to a value of 1.0 in the pattern matrix; the same item was used to scale factors between sexes. Italicized items in Figure 2 were chosen for these scaling purposes. We used the chi-square difference test ($\Delta\chi^2$; Jöreskog & Sörbom, 1989) to compare nested models. The critical value for all comparisons was $p < .01$. Correlated error terms previously described in Nyberg et al. (2003) were modeled for consistency and comparison purposes, and included RN Names with RN Faces; RC Actions with RC Nouns; RC Item with RC Source; all error terms for RC FF, RC DF, RC FD, RC DD; and FluA with FluM. For all comparative models, an education factor was regressed onto all observed variables to control for sex differences in years of education.

Tests of measurement invariance. Hertzog and Nesselroade (2003) provide a conceptual overview of measurement invariance. Tests following the framework of Meredith (1993) included: (1) *Configural invariance*, which examined similar factor patterns. Differences in the magnitude of factor loadings between groups were allowed. (2) *Weak metric invariance of first-order loadings*, which demonstrated no differences between the location and magnitude of factor loadings between sexes. Weak (metric) invariance allows for meaningful quantitative comparisons of groups by establishing comparable measurement units for the variables and factors (Cunningham, 1982, 1991; Horn & McArdle, 1992; Horn, McArdle, & Mason, 1983; Maitland, Dixon, Hultsch, & Hertzog, 2001). (3) *Weak metric invariance of second-order loadings*, in which second-order factor loadings were constrained to be equivalent between sexes. This measure provides information equivalent to (2) for higher order factor models. (4) *Strong invariance of observed variable intercepts*, in which observed variable intercepts are constrained to be equivalent across sexes. This last measure tests invariance of latent means between groups to establish sex differences. Additional tests for *strict invariance*—that is, for whether uniqueness terms are invariant across sexes—were also examined, but they were not supported and are, therefore, not included in our result tables.

Tests of latent mean differences. Sex differences in memory performance were tested using methods described by Byrne (1998) and demonstrated by Maitland and colleagues (Maitland et al., 2000; Nyberg et al., 2003; Schaie et al., 1998). These methods use the τ matrix in LISREL to model the regression intercepts for each observed variable, whereas the α matrix in LISREL is employed to model latent mean deviation values for the first-order factors between groups. An arbitrary reference group is selected (women were selected for all comparisons), and the mean performance for the remaining group(s) is compared with the referent.

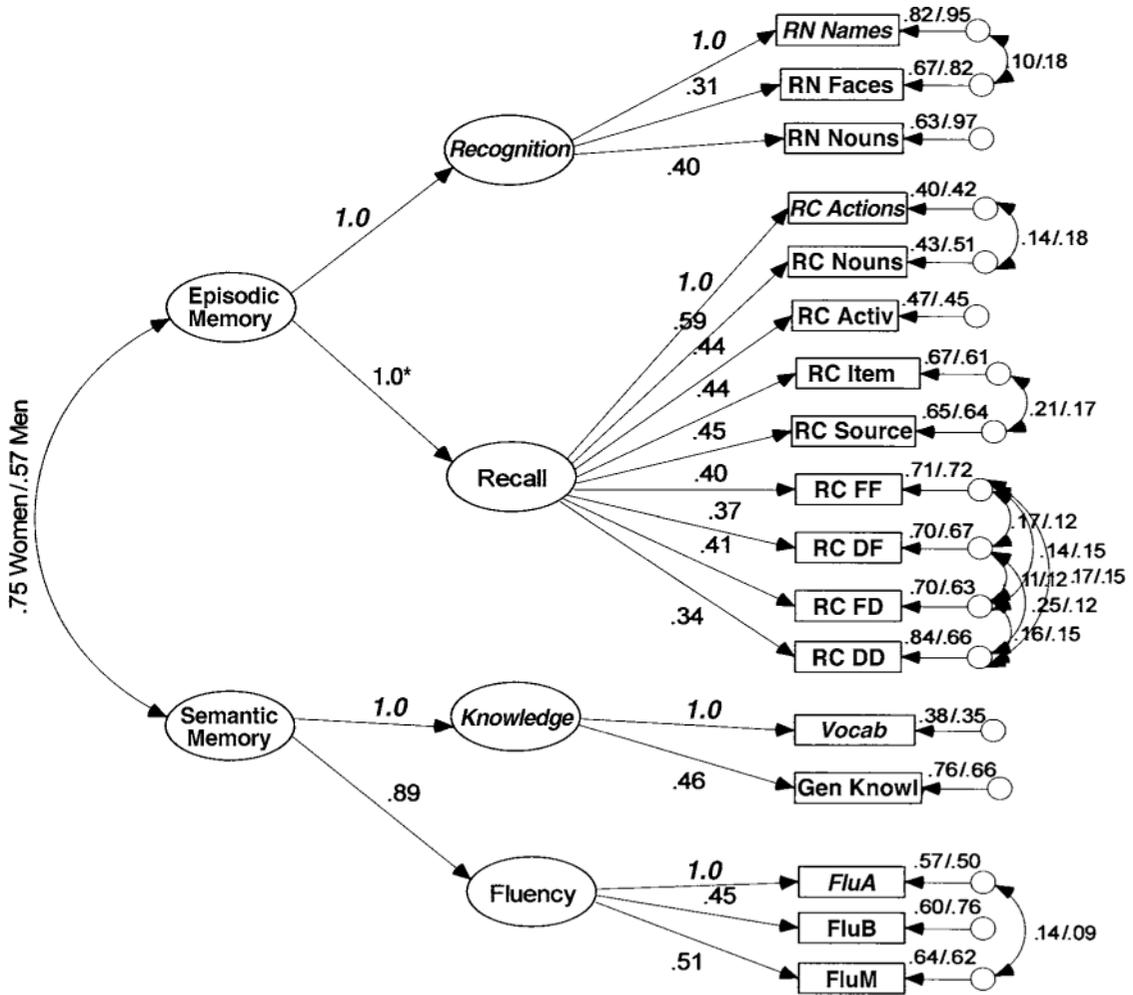


Figure 2. Standardized solution for Model M4 demonstrating weak metric invariance of first- and second-order factor loadings. Italicized items and loadings represent variables used to scale factors, and loadings for women are listed first for each variable. *The factor loading between episodic memory and recall rounded to a value of 1.0.

Two strategies were used to examine first- and second-order factor means. To test latent mean differences for the second-order episodic and semantic memory factors, the variable intercepts for women (reference) were estimated freely. The vector of first-order latent mean values for all groups and the second-order factor means for the reference group were fixed to zero. Intercept values were constrained to be invariant between the sexes (imposing strong invariance), whereas the second-order mean vectors were freely estimated for men. Any deviation value noted in the κ matrix in LISREL that exceeds the critical value ($\pm 2.58, p < .01$) would suggest a significant group difference in second-order factor performance.

To perform the first-order mean comparisons, we fixed values in the α matrix in LISREL to zero for women (reference) for the recognition, recall, knowledge, and fluency factors, whereas intercept values for all observed variables were freely estimated for women (reference). The test of strong invariance required that the intercept values be constrained to be equivalent between sexes. Means for men were also freely estimated, resulting in deviation values from the reference group. All values falling within the confidence band would suggest that no sex differences exist, whereas values falling outside the confidence band would indicate significant sex differences in memory performance.

Comparative fit indexes. Model fit was evaluated by examining the following fit indexes: (1) the model χ^2 ; (2) the goodness-of-fit index (GFI; Jöreskog & Sörbom, 1989); (3) the non-normed fit index (NNFI; Bentler & Bonnett, 1980); (4) the comparative fit index (CFI; Bentler, 1990); (5) the root mean square error of approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980); and (6) the z ratio (Bollen, 1989).

RESULTS

Reliability

Reliability estimates (split-half correlations boosted by the Spearman–Brown formula) ranged from .63 to .74 across lists and conditions in the episodic memory tasks, and from .55 to .64 in the semantic tasks (vocabulary and general knowledge). There were no differences between women and men in these estimates, and the ranges were similar to those reported by Rönnlund, Nyberg, Bäckman, & Nilsson (2003) for another independent sample (Sample 1) of the Betula Project.

Initial analyses revealed no differences between lists and presentation orders for those tasks in which different lists and orders were used in a counterbalanced way. For this reason, the results to be reported were based on collapsed data across these variables.

Appropriateness of Initial Model

Our first model examined whether the six-factor model of declarative memory proposed by Nyberg et al. (2003) fit satisfactorily for women. Results revealed an excellent fit for the model: All factor loadings, factor variances and covariances, and correlated uniqueness terms were statistically significant, allowing us to accept this model as our basic structure to be used in multigroup tests of measurement invariance [M1: $\chi^2(104) = 246.72, p < .001$; GFI = .972, NNFI = .989, CFI = .992, RMSEA = .037]. Results for this model and all nested model comparisons are provided in Table 2. Standardized factor loadings ranged from a low of .49 (RC DD) to a high of .74 (RC Nouns). The correlation between the episodic and semantic memory factors for women was $r = .85$.

Sex Invariance of Declarative Memory

To test the hypothesis of sex invariance of first- and second-order factors, we first added the data for men, expanding our strategy into a multigroup, simultaneous structural model. This initial multigroup model demonstrated evidence for configural invariance of declarative memory in women and men [Configural invariance model M2: $\chi^2(208) = 492.57, p < .001$; GFI_{women} = .975, GFI_{men} = .965, NNFI = .988, CFI = .992, RMSEA = .039]. All factor loadings, factor variances and covariances, and correlated uniqueness terms remained statistically significant, with one exception: The correlated error term FluA and FluM was not statistically significant for men. The correlations between episodic and semantic factors were $r = .73$ and $r = .58$ for women and men, respectively.

Next, we tested whether the first-order factor loadings could be constrained to be equivalent between women and men (a test of weak metric invariance), and the data supported this assumption [Weak metric invariance model M3: $\chi^2(238) = 527.85, p = .001$; GFI_{women} = .973, GFI_{men} = .962, NNFI = .989, CFI = .992, RMSEA = .037; $\Delta\chi^2_{M3-M2}(30) = 35.28, p > .05$]. The same pattern was noted in the correlated uniqueness terms as for the model M2, and the correlations between the episodic and semantic memory factors were $r = .71$ for women and $r = .61$ for men. The next model tested for weak structural invariance of second-order factor loadings between sexes, and this hypothesis was found to be plausible when compared against M3 [Weak metric invariance model M4: $\chi^2(240) = 529.03, p = .001$; GFI_{women} = .973, GFI_{men} = .962, NNFI = .989, CFI = .992, RMSEA = .037; $\Delta\chi^2_{M4-M3}(2) = 1.18, p > .05$]. Thus, the results indicated that the same measurement structure underlies declarative memory performance of both men and women, and this model was accepted as best fitting the data. The results from M4, demonstrating sex invariance for all factor loadings for the entire sample, are presented in Figure 2. The correlations between the episodic and semantic memory factors in the final accepted model were $r = .75$ for women and $r = .57$ for men. Note that separate uniqueness terms are represented for women and men, respectively, as these values could not be constrained to be equivalent. We accept that the six-factor, higher-order model of declarative memory is a parsimonious representation of the data and that substantial measurement invariance exists between the sexes.

Sex Differences in Declarative Memory

After having established support for sex invariance of the factor loadings underlying the structure of declarative memory, the next analyses involved an examination of sex differences in actual declarative memory performance. Although the best-fitting covariance models dem-

Table 2
Comparison of Declarative Memory Models

Initial Model	<i>df</i>	χ^2	<i>p</i>	GFI	RMSEA	NNFI	CFI	<i>z</i> Ratio
Baseline Model								
M1 Women only	104	246.72	.001	.972	.037	.989	.992	2.37
Sex Invariance Models with Education Covaried								
M2 Configural multigroup model	208	492.57	.001	.975 _w .965 _M	.039	.988	.992	2.37
M3 Weak invariance between sexes, first-order factor loadings	238	527.85	.001	.973 _w .962 _M	.037	.989	.992	2.22
M4* Weak invariance between sexes, first- and second-order factor loadings	240	529.03	.001	.973 _w .962 _M	.037	.989	.992	2.20
Models with Latent Means								
M5 Strong invariance between sexes, M4 with first-order means	253	585.85	.001	.973 _w .961 _M	.038	.988	.990	2.31
M6 Strong invariance between sexes, M4 with second-order means	256	616.44	.001	.973 _w .961 _M	.039	.987	.990	2.41

Note—GFI, LISREL goodness-of-fit index; RMSEA, root mean square error of approximation; NNFI, non-normed fit index; CFI, comparative fit index; *z* ratio, χ^2/df . W, Women; M, Men. *M4 accepted as the best-fitting overall sex invariance model.

onstrated measurement and structural equivalence of declarative memory in women and men, this lack of sex differences in the covariance structure of the model does not provide information about sex differences in mean performance for the six declarative memory factors. To address the more global issue concerning sex differences in episodic and semantic memory performance, we employed tests of strong invariance to examine sex differences in the latent means for these second-order factors. Results show that women performed significantly better than men for both episodic and semantic memory when compared across the adult lifespan (see Tables 2 and 3, Figure 3).

Next, we examined sex differences for the four first-order factors (recognition, recall, knowledge, and fluency) as an additional test of strong invariance to determine whether the differences generalized across all subdomains or were more selective. Significant sex differences were noted for both the recognition and recall factors, indicating that the female advantage for episodic memory generalizes across these tasks. Women also performed at a higher level than men for fluency, whereas no significant differences were noted between women and men for knowledge (see Figure 4).

Age-Group-Specific Sex Differences

These results suggest that sex invariance of the measurement model exists across the wide age range within this sample. In addition, previous work has established invariance across three age groups within Betula samples

(Nyberg et al., 2003). Combining Samples 2 and 3 for the current study allows us to expand our upper age range by 5 years compared with Nyberg et al. Therefore, we examined sex invariance within each of three age groups (35–50, 55–65, and 70–85 years) and found strong evidence for sex invariance of all factor loadings for all comparisons. We employed the same sequence of models (M2–M4) described for the overall sex difference comparisons (testing configural invariance [M2], weak invariance of first-order factor loadings [M3], and weak invariance of second-order factor loadings [M4]) and found evidence for excellent model fit and sex invariance of all first- and second-order factor loadings across all three age groups [Middle-aged weak invariance model M4: $\chi^2(240) = 347.95, p = .001$; $GFI_{women} = .949, GFI_{men} = .941, NNFI = .972, CFI = .978, RMSEA = .037$. Young-old weak invariance model M4: $\chi^2(240) = 290.95, p = .014$; $GFI_{women} = .951, GFI_{men} = .939, NNFI = .986, CFI = .989, RMSEA = .027$. Old-old weak invariance model M4: $\chi^2(240) = 369.73, p = .001$; $GFI_{women} = .942, GFI_{men} = .913, NNFI = .976, CFI = .981, RMSEA = .044$]. Whereas strong evidence for sex invariance was found, we acknowledge that the model fit was somewhat stronger for women, regardless of age group.

Informed by these results, we tested strong invariance models to determine whether sex differences at the level of latent means were modified by age. This was done by examining performance differences between sexes separately within each age group. Sex differences were noted

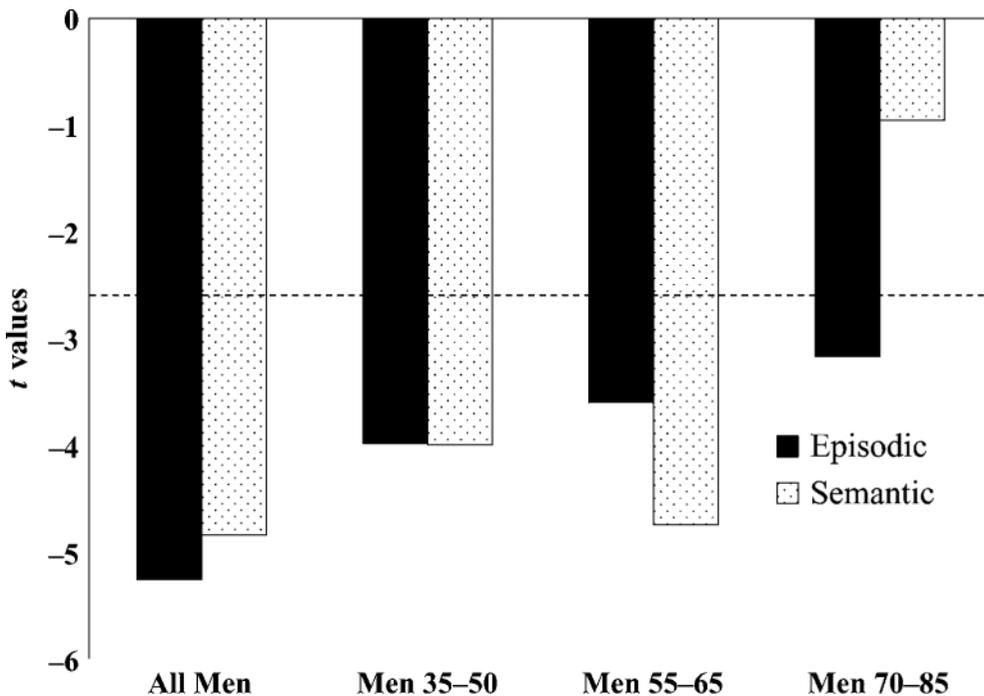


Figure 3. Sex differences for second-order factors. Latent mean values for women were set to zero and served as the reference group to derive deviation values (expressed as *t* values) for men. Negative values indicate lower memory performance for men relative to women. Confidence band is set at 2.58; values exceeding this are statistically significant differences.

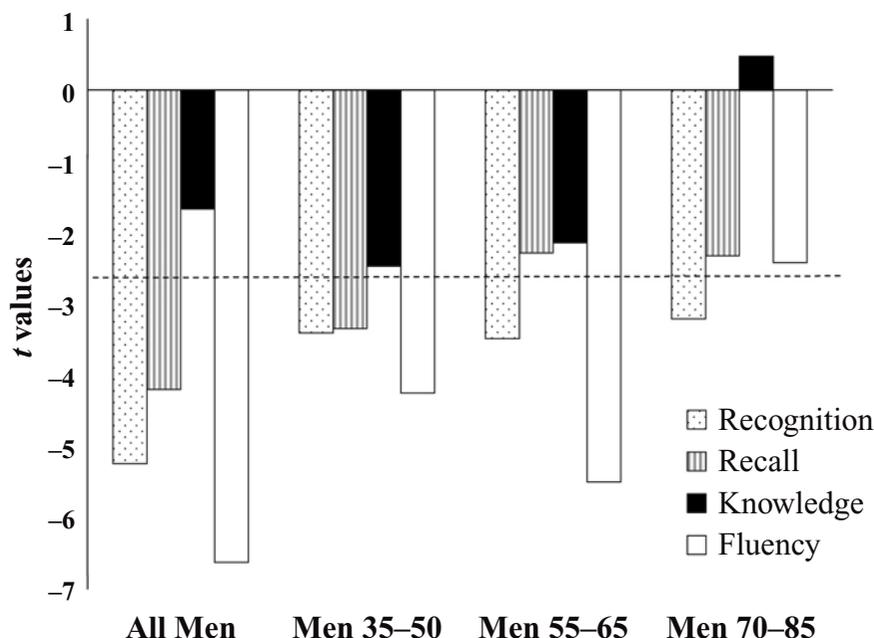


Figure 4. Sex differences for first-order memory factors. See Figure 3 for scaling details.

for the second-order episodic memory factor within each age group, whereas sex differences for semantic memory performance were found only in the middle-aged and young-old groups (see Figure 3). Conversely, we noted sex differences in episodic memory performance for the first-order recognition factor in all three age groups, whereas significant differences favoring women were noted on the recall factor for the middle-aged group only. However, it is important to note that trends in the same direction were observed in both older groups. With regard to semantic memory, no sex differences were noted for the knowledge factor in any age group. Finally, reliable differences in fluency were noted for the middle-aged and young-old groups (see Figure 4).

DISCUSSION

The present results provide strong evidence for the existence of sex differences in declarative memory. Importantly, such differences seem to be selective, in that they are present for some, but not other, declarative memory functions, and many of these sex differences are qualified by age. Thus, the data support a division of declarative memory into episodic and semantic factors, a subdivision of episodic memory into recognition and recall factors, and a subdivision of semantic memory into knowledge and fluency factors, as described by Nyberg et al. (2003). Furthermore, the present study provides evidence for equivalence in the organization and structure of declarative memory for women and men across much of the adult lifespan. This is an important finding, because group comparisons on any given task depend on the implicit assumption that the underlying structures do not

differ between the groups in question. Whereas comparisons of women and men on various cognitive measures are quite common, studies testing the implied structure of cognition for measurement equivalence across sexes are rare. Instead, researchers assume measurement invariance without having first completed the appropriate comparative analyses (Vandenberg & Lance, 2000). Regarding declarative memory, some studies have examined the structure of memory without formal tests of measurement invariance (e.g., Allen, Sliwinski, Bowie, & Madden, 2002). In addition, some formal tests of sex invariance of factor models do exist, but they have often examined only a single memory factor in relation to other cognitive factors (e.g., Maitland et al., 2000).

Although the structures of declarative memory were equivalent for both sexes, we found that women and men did perform differently on tasks assessing episodic and semantic memory. Specifically, women performed at a higher level than men on recall and recognition of episodic information. Furthermore, women performed at a higher level than men on verbal fluency, whereas no differences were noted in knowledge. Thus, the pattern of sex differences observed at the level of latent means is in agreement with what has been reported at the manifest level in prior research (e.g., Herlitz et al., 1997; Hyde & Linn, 1988).

The fact that the female advantage in episodic memory was not reduced in recognition relative to recall suggests that differences in encoding rather than retrieval underlie the female superiority in episodic memory. Had retrieval problems been prominent in men, one would have expected an attenuation of sex differences in recognition relative to recall. These data also do not support the view that women's greater verbal production skills are critical

Table 3
Mean Deviation Values, Standard Errors, and *t* Values for Sex Comparisons

Factor	All Men			Middle-Aged Men			Young-Old Men			Old-Old Men		
	α	SE	<i>t</i>	α	SE	<i>t</i>	α	SE	<i>t</i>	α	SE	<i>t</i>
Recognition	-0.288	0.055	-5.238	-0.332	0.098	-3.404	-0.361	0.104	-3.484	-0.297	0.093	-3.203
Recall	-0.591	0.141	-4.197	-0.697	0.209	-3.341	-0.747	0.327	-2.281	-0.709	0.306	-2.320
Knowledge	-0.315	0.190	-1.663	-0.635	0.258	-2.466	-0.717	0.336	-2.137	0.194	0.411	0.472
Fluency	-1.085	0.164	-6.619	-1.254	0.296	-4.244	-1.542	0.281	-5.495	-0.740	0.306	-2.415
Episodic memory	-0.193	0.037	-5.253	-0.298	0.075	-3.979	-0.244	0.068	-3.595	-0.183	0.058	-3.165
Semantic memory	-0.793	0.164	-4.837	-0.892	0.223	-3.993	-1.319	0.278	-4.740	-0.353	0.372	-0.949

Note—*t* values less than ± 2.58 are not statistically significant and indicate no sex differences of the latent construct. Positive values indicate better performance, negative values indicate worse performance than the reference group (women).

to their higher episodic memory proficiency, since the demands on self-initiated processing are greater in recall than in recognition (e.g., Craik & McDowd, 1987).

We also examined whether the magnitude of sex differences changes as a function of age. In general, our analyses showed that the size of the female superiority decreased with advancing age, an effect that was most readily observed at the first-order level. Specifically, a female advantage was observed for the episodic factor across all three age groups, and for the semantic factor such an advantage was noted for the younger two age groups. Furthermore, in middle-aged adults, the episodic advantage for women was found for both recognition and recall; in the two older age groups, sex differences were reliable for recognition only. The sex difference in semantic memory was found only for fluency and was restricted to participants aged 35–65 years.

Although our division of participants into the present age groups does not permit an adequate test of the influence of endocrinological factors on the results obtained (e.g., Liben et al., 2002; Sherwin, 1997), a hormonal explanation of sex differences in declarative memory receives some support from the present data. Such an explanation assumes that women with higher levels of estrogen will perform at a higher level than women with lower levels of estrogen (Kampen & Sherwin, 1994; Sherwin, 1997). Following this reasoning, we would expect larger sex differences in the middle-aged group than in the older age groups.

Our finding of sex differences favoring women on recall, recognition, and fluency but not on knowledge provides some clues as to the basis for these differences. One factor that seems to differentiate recall, recognition, and fluency from knowledge is that the former forms of memory are dependent on the hippocampus and related structures in the medial temporal lobe (MTL), whereas this does not seem to be true for knowledge. Specifically, successful encoding of information into episodic memory is known to be dependent on MTL regions (e.g., Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998). Much evidence also shows that MTL regions are engaged during episodic memory retrieval (e.g., Schacter & Wagner, 1999) and that the hippocampal region is important for both recognition and recall (Squire & Knowl-

ton, 2000). Fluency has most strongly been associated with frontal lobe functioning, but recent fMRI findings show that this form of memory is also associated with the hippocampus and the parahippocampal gyrus (Pihlajamäki et al., 2000). In contrast, retrieval of semantic knowledge appears to engage lateral parts of temporal and parietal cortex rather than the hippocampal region (Thompson-Schill, 2003). When we take all of the present findings together, one possible interpretation is that sex differences favoring women are present on memory tests that depend on the hippocampus and related areas of the MTL. This interpretation receives support from findings that the hippocampus is the location of a number of sex differences, including both anatomical differences in the pattern of dendritic branching and differences in sensitivity to environmental enrichment (see McEwen, 2000).

REFERENCES

- ALLEN, P. A., SLIWINSKI, M., BOWIE, T., & MADDEN, D. J. (2002). Differential age effects in semantic and episodic memory. *Journals of Gerontology: Psychological Sciences & Social Sciences*, *57*, P173-P186.
- ALLÉN, S. (1972). *Tiotusen i topp* [Top ten thousand]. Stockholm: Almqvist & Wiksell.
- ANSTAY, K. J., HOFER, S. M., & LUSZCZ, M. A. (2003). Cross-sectional and longitudinal patterns of dedifferentiation in late-life cognitive and sensory function: The effects of age, ability, attrition, and occasion of measurement. *Journal of Experimental Psychology: General*, *132*, 470-487.
- BADDELEY, A. D., LEWIS, V., ELDRIDGE, M., & THOMSON, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*, 518-540.
- BALTES, P. B., REESE, H. W., & NESSELROADE, J. R. (1988). *Life-span developmental methodology: Introduction to research methods*. Hillsdale, NJ: Erlbaum.
- BENTLER, P. M. (1990). Fit indices, LaGrange multipliers, constraint changes, and incomplete data in structural models. *Multivariate Behavioral Research*, *25*, 163-172.
- BENTLER, P. M., & BONNETT, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-600.
- BOLLEN, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- BREWER, J. B., ZHAO, Z., DESMOND, J. E., GLOVER, G. H., & GABRIELI, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, *281*, 1185-1187.
- BYRNE, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

- CRAIK, F. I. M., & MCDOWD, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 474-479.
- CUNNINGHAM, W. R. (1982). Factorial invariance: A methodological issue in the study of psychological development. *Experimental Aging Research*, **8**, 61-65.
- CUNNINGHAM, W. R. (1991). Issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 107-113). Washington, DC: American Psychological Association.
- DUREMAN, I. (1960). *SRB:1*. Stockholm: Psykeförlaget.
- FEINGOLD, A. (1996). Cognitive gender differences: Where are they and why are they there? *Learning & Individual Differences*, **8**, 25-32.
- FOLSTEIN, M. F., FOLSTEIN, S. E., & MCHUGH, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189-198.
- HALPERN, D. F., & LAMAY, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, **12**, 229-246.
- HERLITZ, A., AIRAKSINEN, E., & NORDSTRÖM, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology*, **13**, 590-597.
- HERLITZ, A., NILSSON, L.-G., & BÄCKMAN, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, **25**, 801-811.
- HERTZOG, C., & NESSELROADE, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology & Aging*, **18**, 639-657.
- HORN, J. L., & MCARDLE, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, **18**, 117-144.
- HORN, J. L., MCARDLE, J. J., & MASON, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist*, **1**, 179-188.
- HYDE, J. S., FENNEMA, E., & LAMON, S. J. (1990). Gender differences in mathematics performance: A metaanalysis. *Psychological Bulletin*, **107**, 139-155.
- HYDE, J. S., & LINN, M. C. (1988). Gender differences in verbal ability: A metaanalysis. *Psychological Bulletin*, **104**, 53-69.
- JÖRESKOG, K. G., & SÖRBOM, D. (1989). LISREL 7: A guide to the program and applications. Chicago: SPSS.
- JÖRESKOG, K. G., & SÖRBOM, D. (2001). LISREL 8 (Version 8.51) [Computer software]. Chicago: Scientific Software International.
- KAMPEN, D. L., & SHERWIN, B. B. (1994). Estrogen use and verbal memory in healthy postmenopausal women. *Obstetrics & Gynecology*, **83**, 979-983.
- KRAMER, J. H., DELIS, D. C., KAPLAN, E., O'DONNELL, L., & PRIFTERA, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology*, **4**, 577-584.
- LABOVIE, E. W. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 493-502). Washington, DC: American Psychological Association.
- LIBEN, L. S., SUSMAN, E. J., FINKELSTEIN, J. W., CHINCHILLI, V. M., KUNSELMAN, S., SCHWAB, J., DUBAS, J. S., DEMERS, L. M., LOOKINGBILL, G., D'ARCANGELO, M. R., KROGH, H. R., & KULIN, H. E. (2002). The effects of sex steroids on spatial performance: A review and an experimental clinical investigation. *Developmental Psychology*, **38**, 236-253.
- MAITLAND, S. B., DIXON, R. A., HULTSCH, D. F., & HERTZOG, C. (2001). Well-being as a moving target: Measurement equivalence of the Bradburn Affect Balance Scale. *Journals of Gerontology: Psychological Sciences & Social Sciences*, **56**, P69-P77.
- MAITLAND, S. B., INTRIERI, R. C., SCHAIE, K. W., & WILLIS, S. L. (2000). Gender differences and changes in cognitive abilities across the adult life span. *Aging, Neuropsychology, & Cognition*, **7**, 32-53.
- MC EWEN, B. S. (2000). Stress, sex, and the structural and functional plasticity of the hippocampus. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 171-197). Cambridge, MA: MIT Press.
- MEREDITH, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, **58**, 525-543.
- NILSSON, L.-G. (1999). Aging, dementia, and memory. In L.-G. Nilsson and H. J. Markowitsch (Eds.), *Cognitive neuroscience of memory* (pp. 147-162). Göttingen: Hogrefe & Huber.
- NILSSON, L.-G., BÄCKMAN, L., ERNGRUND, K., NYBERG, L., ADOLFSSON, R., BUCHT, G., KARLSSON, S., WIDING, M., & WINBLAD, B. (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychology, & Cognition*, **4**, 1-32.
- NYBERG, L., MAITLAND, S. B., RÖNNLUND, M., BÄCKMAN, L., DIXON, R. A., WAHLIN, Å., & NILSSON, L.-G. (2003). Selective adult age differences in an age-invariant multi-factor model of declarative memory. *Psychology & Aging*, **18**, 149-160.
- PIHLAJAMÄKI, M., TANILA, H., HÄNNINEN, T., KÖNÖNEN, M., LAAKSO, M., PARTANEN, K., SOININEN, H., & ARONEN, H. J. (2000). Verbal fluency activates the left medial temporal lobe: A functional magnetic resonance imaging study. *Annals of Neurology*, **47**, 470-476.
- RÖNNLUND, M., NYBERG, L., BÄCKMAN, L., & NILSSON, L.-G. (2003). Recall of subject-performed tasks, and cognitive activities across the adult life span: Parallel age-related deficits. *Aging, Neuropsychology, & Cognition*, **10**, 182-201.
- ROSÉN, M. (1995). Gender differences in structure, means and variances of hierarchically ordered ability dimensions. *Learning & Instruction*, **5**, 37-62.
- SCHACTER, D. L., & WAGNER, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies on episodic encoding and retrieval. *Hippocampus*, **9**, 7-24.
- SCHAIE, K. W., MAITLAND, S. B., WILLIS, S. L., & INTRIERI, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology & Aging*, **13**, 8-20.
- SHERWIN, B. B. (1997). Estrogen effects on cognition in menopausal women. *Neurology*, **48** (Suppl. 7), 21-26.
- SMALL, B. J., VIITANEN, M., & BÄCKMAN, L. (1997). Mini-Mental State Examination item scores as predictors of Alzheimer's disease: Incidence data from the Kungsholmen project, Stockholm. *Journals of Gerontology: Biological Sciences & Medical Sciences*, **52**, M299-M304.
- SQUIRE, L. R., & KNOWLTON, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 765-779). Cambridge, MA: MIT Press.
- STEIGER, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173-180.
- STEIGER, J. H., & LIND, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, IA.
- THOMPSON-SCHILL, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia*, **41**, 280-292.
- VANDENBERG, R. J., & LANCE, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, **2**, 4-69.
- VOYER, D., VOYER, S., & BRYDEN, M. P. (1995). Magnitude of sex differences in spatial abilities: A metaanalysis and consideration of critical variables. *Psychological Bulletin*, **117**, 250-270.
- WAGNER, A. D., SCHACTER, D. L., ROTTE, M., KOUTSAAL, W., MARIL, A., DALE, A. M., ROSEN, B. R., & BUCKNER, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, **281**, 1188-1191.