Aging, Neuropsychology, and Cognition
2000, Vol. 7, No. 1, pp. 32-53

1382-5585/00/0701-032$15.00

Gender Differences and Changes in Cognitive Abilities Across the Adult Life Span*

Scott B. Maitland[1], Robert C. Intrieri[2], K. Warner Schaie[3], and Sherry L. Willis[3]
[1]University of Guelph, [2]Western Illinois University, and [3]The Pennsylvania State University

ABSTRACT

Gender differences in the covariance structure and latent means of cognitive abilities across the adult life span are explored. A multiply marked, six-factor measurement model examining 982 participants from the Seattle Longitudinal Study tested gender and longitudinal invariance over 7 years. Substantive invariance tests on the factor correlations were also examined. Longitudinal results for the entire sample indicated the best fit was a partial invariance solution supporting time and gender invariance. Individual analyses showed younger and middle age groups to be metrically invariant for all factor loadings as well as time and gender invariance of factor variances and intercorrelations. A metric model also fit best for the older group, although factor variances and intercorrelations only demonstrated gender invariance. Latent mean structures revealed gender differences with women outperforming men for Verbal Recall, and in younger and middle age groups Perceptual Speed, whereas men consistently outperformed women for Spatial Orientation. Results demonstrate gender equivalence in covariance structures across the adult life span and within separate age groups, and reveal gender differences and changes over 7 years in cognitive abilities at the latent mean level.

Gender and age-related differences in intellectual development have intrigued psychologists for years. Numerous studies document gender and age differences in the profile of cognitive abilities, especially among children and adolescents (e.g., Hyde & Linn, 1988; Linn & Petersen, 1985, 1986; Maccoby & Jacklin, 1974). However, systematic attention to gender differences in patterns of intellectual abilities across adulthood remains sparse. The study of gender differences in cognitive abilities should include an examination of the covariance structure of the abilities under consideration and mean level performance differences between men and women. Most research to date has examined the latter while ignoring the former.

Schaie and Hertzog (1983) reported consistent gender differences in observed means for primary mental ability (PMA) performance. Men outperformed women on space and number factors, whereas women did better than men on word fluency. The strong sex difference on spatial ability is consistent with results found in younger populations and corresponds with Schaie's earlier findings (Strother, Schaie, & Horst, 1957) and those of others (Deaux, 1985; Maccoby & Jacklin, 1974). More specifically, meta-analyses of spatial ability tasks have shown larger gender differences favoring men for mental rotation tasks than for spatial perception or spatial visualization tasks (Linn & Peterson, 1985; Voyer, Voyer, & Bryden, 1995).

---

* The research reported in this article was supported by Grant R37 AG4770 from the National Institute on Aging. We gratefully acknowledge the enthusiastic cooperation of members and staff of the Group Health Cooperative of Puget Sound.
Address correspondence to: Scott B. Maitland, Department of Family Relations and Applied Nutrition, University of Guelph, Guelph, ON, Canada N1G 2W1. Tel.: (519) 824-4120, ext. 6156. Fax: (519) 766-0691. E-mail: smaitlan@uoguelph.ca.
Accepted for publication: November 23, 1999.

Additionally, research has stressed the *diminishing effect* of gender differences and the shrinking of the *gender gap* in other intellectual skills (Feingold, 1988; Hyde & Linn, 1988). Even so, some gender-related intellectual differences in level of performance have remained (Feingold, 1988; Kaufman, Kaufman-Packer, McLean, & Reynolds, 1991). Moreover, concern has focused on the shrinking cohort and gender differences in mean performance whereas little attention has been directed toward changes in the gender-specific latent structure of cognitive abilities across the adult life span.

Hertzog and Carter (1982) systematically examined aging, gender, and cognitive abilities using a structural equation framework. Four generations were examined for gender differences. Isolating verbal and spatial intelligence factors revealed substantial invariance in the two-factor solution and supports similar intellectual factor structures for men and women. Hertzog and Carter (1982) noted group differences in factor means, with men outperforming women on the spatial ability factor whereas no mean level gender differences were observed on the verbal performance factor. As a result, they concluded that their results provided no support for the differentiation of gender-related organization of cognitive abilities.

Kaufman et al. (1991) compared men and women on the Wechsler Adult Intelligence Scale – Revised (WAIS-R) to determine differences in crystallized and fluid abilities. Covarying by level of education, the authors concluded that the patterns for fluid and crystallized intelligence are similar for men and women. However, men outperformed women on arithmetic for all age groups. Consistent and substantial WAIS-R gender differences were noted on Information, Block Design, and Digit Symbol. Digit Symbol results clearly favored women whereas men generally performed better on Information and Block Design. Both men and women maintained their crystallized abilities through old age but demonstrated rapid early decline in fluid ability.

Support for women's superior performance in verbal abilities over men has been widely reported (e.g., Halpern, 1992; Hart & O'Shanick, 1993; Hyde & Linn, 1988). Furthermore, studies have demonstrated gender differences favoring women in word recall as a measure of episodic memory (e.g., Schaie & Willis, 1993; Zelinski, Gilewski, & Schaie, 1993). Herlitz, Nilsson, and Bäckman (1997) noted that women's higher performance in episodic memory tasks may be related to higher verbal abilities because reports of memory performance employ verbal tasks. They reported gender differences in episodic memory favoring women despite controlling for verbal fluency. Whereas they found gender differences in episodic memory, no differences between genders were found for measures of semantic memory, primary memory, or priming.

Recently, a meta-analysis of 25 studies supported gender differences favoring women on verbal episodic memory tasks, a nonsignificant female advantage for verbal fluency, and male advantages on spatial, primary working/working memory, and tests of reasoning (Meinz & Salthouse, 1998). The magnitude of gender differences was quite small; however, spatial abilities demonstrated the largest divergence between genders.

Hultsch, Hertzog, Dixon, and Small (1998) examined gender differences in the Victoria Longitudinal Study (VLS) from a variety of different perspectives. Cross-sectional analyses showed men performed better on semantic memory measures and a higher order crystallized intelligence factor, whereas women performed better on episodic memory tasks. However, whereas regressing age, gender, and education onto cognitive factors in longitudinal models demonstrated individual differences in the amount of cognitive change occurring in later adulthood, little evidence was found for gender differences in the rate of change over a 6-year period. Similarly, Zelinski and Burnight (1997) reported no differential patterns of change in memory as a function of gender when examining 16-year retest data from the Long Beach Longitudinal Study.

Finally, a study of gender differences of observed scores, latent structures, means, and variances of cognitive ability dimensions in a sample of adolescents provides a segue to the current study. Whereas Rosén (1995) does not discuss modeling approaches in the terminology of factorial invariance put forth by Meredith and others, equivalence of covariance structures and latent mean gender differences were compared. The pattern of gender differences for latent means was quite discrepant from observed, single-indicator, test scores. This finding provides further support for the importance of examining gender differences with multiply-marked, latent factor analytic techniques.

The current study focused on the equivalence of the covariance structure of adult cognitive abilities over the adult life span, and gender differences and change in mean level performance of these latent abilities. A five-factor measurement model of cognitive abilities was established by Schaie, Willis, Jay, and Chipuer (1989) and later modified to include a verbal recall factor (Schaie, Dutta, & Willis, 1991). This model was recently tested for age-group and time invariance using six age/cohort groups across a 7-year interval (Schaie, Maitland, Willis, & Intrieri, 1998). The six factors include: (a) Inductive Reasoning, (b) Spatial Orientation, (c) Verbal Comprehension, (d) Numerical Facility, (e) Perceptual Speed, and (f) Verbal Recall. To understand the similarities and differences between men and women, a confirmatory factor analysis with nonzero latent means (Everson, Millsap, & Rodriguez, 1991; Jöreskog & Sörbom, 1979; Millsap & Everson, 1991) was used to examine the cognitive abilities model.

Gender differences in the factor structures suggest that the cognitive ability factors correlate differentially for men and women. Differences in the latent or common factor means indicate the presence of ability-specific differences between gender groups. Determining ability-specific gender differences rather than comparisons of global intelligence was an important component of the current study. Emphasis on comparisons of individual latent factor mean differences will help us to better understand ability-specific differences between men and women, and in separate age groups.

First, models examining 7-year longitudinal gender comparisons are provided. Second, models testing the covariance structure for gender differences and change within three age groups are presented. Finally, accepted covariance structures were extended and latent factor means compared between gender groups. Latent factor mean changes were examined longitudinally and between age and gender groups.

The following questions are considered: (a) How well does the six-factor measurement model of cognitive abilities described by Schaie and colleagues (1991, 1998) fit for men and women? (b) What level of invariance does the covariance structure fit best? (c) Do men and women significantly differ in cognitive abilities at the latent factor mean level? (d) Do gender differences exist in the patterns of change in cognitive abilities over 7 years? (e) What gender differences exist, both in the covariance structure and for latent factor means, when younger, middle, and older age groups of adults are examined?

METHOD

Participants

The Seattle Longitudinal Study (SLS) has examined adult cognitive functioning for approximately 40 years. Randomly sampled participants, who were members of a health maintenance organization in the Pacific Northwest and equally distributed by gender and age, were tested at 7-year intervals (Schaie, 1996). Sample survivors were retested and additional panels added at each study phase. The sampling frame represents broadly distributed educational and occupational levels, covering the upper 75% of the socioeconomic spectrum.

These data include 982 community-dwelling individuals (442 men and 540 women) examined in both the fifth SLS cycle (1983 – 1985) and the sixth cycle (1990 – 1992) with an average time interval of approximately 7 years. Average educational level was 14.7 years ($SD = 3.0$; Range: 1 – 20 years) and mean family income was $27,884 ($SD$ = $6,484; Range: $1,000 – $50,000+). Occupations were rated on a scale from 0 for unskilled to 9 for professional occupations. Mean occupational level for gainfully employed individuals was 6.9 ($SD = 1.87$; Range: 0 – 9). Most frequent occupations involved skilled trades, clerical sales, and managerial and semiprofessional jobs (see Schaie, 1996).

The sample was divided into three age groups defined as younger, middle, and older adults with mean ages at time of first measurement of 35.5

(range 22 – 49), 56.5 (range 50 – 63), and 75.5 (range 64 – 87) years respectively. Thus, there was approximately 20 years difference in the mean age between groups. There were 296 (men = 134, women = 162) younger adults, 330 (men = 154, women = 176) middle adults, and 356 (men = 154, women = 202) older adults. Younger men had 15.9 years of education, slightly greater than one year more than younger women (14.8 years) had. Middle age group men had 15.2 years of schooling compared to 14.2 years for women in this age group. Finally, older men had 14.3 years of education whereas older women had 13.7 years. Similarly, men reported higher income than women for all three age groups (younger: $31,630 vs. $30,276; middle: $29,840 vs. $28,154; older: $25,540 vs. $21,868 respectively).

Measurement Variables

The original SLS battery administered to all study waves includes the five PMA measures. The battery was expanded in 1984 to permit structural analyses that require multiple measures to mark each cognitive ability factor. The measures, the primary ability that they mark, their sources, and their test-retest correlations over a 2-week interval for a group of 172 participants have been reported elsewhere (Schaie et al., 1998; Schaie, Willis, Hertzog, & Schulenberg, 1987). The test-retest correlations ranged from .73 to .95, indicating satisfactory reliability for all instruments. Under the assumption of perfect stability of individual differences in the true scores, these correlations estimate the reliability of the tests (Schaie & Hertzog, 1985). To the extent that individual differences are not perfectly stable, these correlations underestimate the marker's reliability. A brief description of the primary abilities and the measures marking them is given below.

Inductive Reasoning

PMA Reasoning (Thurstone & Thurstone, 1949) assesses inductive reasoning ability via letter series problems. Participants are shown a series of letters and must select the next letter in the series from 5 letter choices. The Adult Development and Enrichment Project Test (ADEPT) Letter Series test (Blieszner, Willis, & Baltes, 1981) also contains letter series problems; however, some of the problems involve pattern description rules other than those found on the PMA measure. The Word Series test (Schaie, 1985) parallels the PMA measure in that the same pattern description rule is used for each item; however, the test stimuli are days of the week or months of the year, rather than letters. The Number Series test (Ekstrom, French, Harman, & Derman, 1976) involves series of numbers and different types of pattern description rules involving mathematical computations.

Spatial Orientation

Three of these tests (PMA Space, Object Rotation, Alphanumeric Rotation) are multiple response measures of two-dimensional mental rotation ability. Participants are shown a model line drawing and asked to identify which of six choices shows the model drawn in different spatial orientations. Stimuli for the PMA Space test are abstract figures. The Object Rotation test (Schaie, 1985) and the Alphanumeric Rotation test (Willis & Schaie, 1983) were constructed so the angle of rotation in each answer choice is identical with the angle used in the PMA Spatial Orientation test (Thurstone & Thurstone, 1949). Tests vary in item content. Object Rotation involves drawings of familiar objects and the Alphanumeric test contains letters and numbers. Cube Comparison (Ekstrom et al., 1976) requires the matching of degree of rotation for three-dimensional cubes upon mental rotation.

Verbal Comprehension

All measures are multiple choice tests that require selecting a synonym for a stimulus word from four alternatives. Tests include the PMA Verbal Meaning test (Thurstone & Thurstone, 1949), and levels 2 and 4 respectively from the Educational Testing Services (ETS) Factor Reference Kit (Ekstrom et al., 1976).

Numerical Facility

The PMA Number test involves checking simple addition problems (Thurstone & Thurstone, 1949). The Addition test (Ekstrom et al., 1976) involves checking the accuracy of numerical summation of problems. The Subtraction and Multiplication test (Ekstrom et. al., 1976), requires calculating the sums and products for alternate rows of simple subtraction and multiplication problems.

Perceptual Speed

All perceptual speed measures come from the ETS Factor Reference Kit (Ekstrom et al., 1976). Finding As involves the cancellation of the letter *a* in columns of words. Picture Identification requires finding the match among five simple test figures to a stimulus figure. Number Comparison involves comparing two sets of eight digit numbers and marking those pairs that are not identical.

Verbal Recall

The Immediate Recall task requires participants to study a list of 20 words for 3.5 minutes. They are then given an equal period of time to recall the words in any order. Delayed Recall requires recalling the Immediate Recall words after an hour of continued testing.

Statistical Analysis Procedures

Models were tested using LISREL 8 (Jöreskog & Sörbom, 1996). Factor structure hypotheses were tested at three levels of stringency: (a) *configural invariance* permitted similar factor patterns whereas differences in factor loadings (lambda) and factor variances and covariances (psi) between groups and over time were allowed; (b) *metric invariance*, implying no difference between the factor loadings and pattern (lambda) constrained between groups, over time, or constrained simultaneously between groups and over time; and (c) *invariance of factor intercorrelations*, with no differences between the factor loadings and patterns (lambda), and constraining the factor variances and covariances (psi) between groups, over time, or both. Table 1 provides a schematic diagram of the appropriate model constraints to test these models. Elements of the factor loading (LY) and factor intercorrelation (PSI) matrices are either allowed to be freely estimated (FR) or constrained to be invariant (IN) across gender, time, or both gender and time as prescribed for each model.

Additionally, if metric invariance of all factor loadings was not accepted, tests of *partial* measurement invariance were employed. Partial invariance implies that some, but not all, of the configurally invariant items have equivalent metric factor loadings (Byrne, Shavelson, & Muthén, 1989). Tests of partial measurement invariance were conducted on the individual markers (factor loadings) upon rejection of the hypothesis positing all factor loadings to be equivalent. Items that were demonstrated to be equivalent were constrained across groups or over time, whereas loadings which were significantly different between groups or over time were allowed to vary freely. This approach allows researchers to identify which items have variability in measurement properties and to adjust statistical analyses for these differences (see Byrne et al., 1989; Schaie et al., 1998).

The current analyses tested the six-factor cognitive abilities model as a repeated measures factor model for panel data (Alwin, 1988; Jöreskog, 1979). Analyses used the variance-covariance matrix, with results of the final models reported as standardized estimates for ease of interpretation. Models were scaled by fixing the best marker for each factor to 1.0 in the pattern matrix (LY). The same item was used to scale each factor for both genders and across time. A 40 × 40 matrix of Time 1 (1984) and Time 2 (1991) scores, containing the same 20 cognitive measures comprising six factors for each occasion, was used to examine longitudinal analyses. The longitudinal model accounted for autocorrelation of the residuals because repeated measures of the same scales were used (Sörbom, 1975; Wiley & Wiley, 1970). Additionally, once the level of factorial invariance was established, latent mean structures were examined to determine gender differences and change over 7 years. Methods described by Byrne (1998) and Schaie et al. (1998) were employed to test these relationships.

Table 1. Comparison of Model Constraints for Longitudinal Analyses.

	Factor Loadings (LY)		Factor Variances – Covariances (PSI)	
Model	Gender	Time	Gender	Time
Configural invariance				
CF	FR	FR	FR	FR
Metric invariance				
M1	FR	IN	FR	FR
M2	IN	FR	FR	FR
M3	IN	IN	FR	FR
Invariance of factor variances and correlations				
PSI1	IN	IN	FR	IN
PSI2	IN	IN	IN	FR
PSI3	IN	IN	IN	IN

Note. M1 and M2, and PSI1 and PSI2, are not nested models. LY = Lambda-Y matrix of factor loadings in LISREL; PSI = PSI matrix of factor variances and covariances in LISREL; CF = configural invariance model; M1-M3 = metric invariance models; PSI1-PSI3 = invariance of factor variances and correlations models; IN = invariance constraint; FR = freely estimated.

Comparative fit statistics
Model fit was evaluated by seeking consensus across an array of multiple fit criteria: (a) χ^2; chi-square is sensitive to sample size and departures from normality (James, Muliak, & Brett, 1982); (b) Goodness of Fit Index (GFI; Jöreskog & Sörbom, 1989); (c) Non-Normed Fit Index (NNFI; Bentler & Bonnett, 1980); (d) Comparative Fit Index (CFI; Bentler, 1990); (e) Root Mean Square Error of Approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980); and (f) Z-Ratio (Bollen, 1989). Incremental fit indexes supply information of practical significance in which statistically significant effects may be evaluated for their usefulness in explaining the data.

The NNFI is useful in confirmatory factor studies to compare the fit of a proposed model to the fit of a null model where all the variables are assumed to be uncorrelated. Values less than .9 indicate that the model may be substantially improved. The CFI is a robust measure of comparative fit with values ranging from 0 – 1.0, with values closer to 1.0 suggesting a better fitting model. The GFI is a measure of the relative amount of variances and covariances accounted for by the model (Jöreskog & Sörbom, 1989). RMSEA provides a measure of discrepancy per degree of freedom of the model. The lower limit of RMSEA is 0 and is equivalent to a perfect fitting model. Values less than or equal to .08 are considered a good fit of the model relative to its degrees of freedom (Browne & Cudek, 1993). Finally, the Z-Ratio assesses how large the chi-square estimate is compared to the expected value (the degrees of freedom; Bollen, 1989). Opinions on what represent "good" fit range from ratios of 3 or lower (Carmines & McIver, 1981) to a high of 5.

The chi-square difference test ($\Delta\chi^2$; Jöreskog & Sörbom, 1989) was used as a basis for comparing nested models. If the $\Delta\chi^2$ was large compared to the degrees of freedom and evaluated against the critical value of chi-square, the null hypothesis of no significant differences between models was rejected. The critical value used for all comparisons was $p < .01$.

RESULTS

Longitudinal Measurement Models

Longitudinal models tested 7-year age-related change and longitudinal change in the pattern of gender differences in the covariance structures. To test for metric invariance of factor loadings, three models were considered. The first tested the hypothesis of time invariance of factor loadings. Next, gender invariance of the factor loadings was tested. Finally, a simultaneous test of invariance between genders and across time was examined. See Table 2 for comparative fit indexes for these models.

The configural model (L1) allowed factor loadings to be freely estimated for both gender and occasion ($\chi^2 = 2328.81$, $df = 1296$, $p < .001$, GFI = .894, CFI = .975). Model L2 tested time invariance, constraining factor loadings across occasion and estimating loadings separately for men and women ($\chi^2 = 2371.03$, $df = 1330$, $p < .001$, GFI = .892, CFI = .975; $\Delta\chi^2 = 42.21$, $df = 34$, $p > .16$). Comparing the time invariance model to the configural model allowed us to ac-

Table 2. Comparison of Longitudinal (L) Gender Models.

Model	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	Z-Ratio
L1 Configural	1296	2328.81	.001	.89	.029	.97	.98	1.80
L2 Time invariant LY	1330	2371.03	.001	.89	.028	.97	.98	1.78
L3 Gender invariant LY	1330	2396.18	.001	.89	.029	.97	.97	1.80
L4 Partial gender invariant LY	1328	2373.13	.001	.89	.028	.97	.97	1.79
L5 L4 + time invariant LY[a]	1346	2390.97	.001	.89	.028	.97	.97	1.78
L6 Time invariant PSI	1388	2627.03	.001	.88	.030	.97	.97	1.89
L7 Gender invariant PSI[b]	1424	2478.83	.001	.89	.028	.97	.97	1.74
L8 Metric invariant PSI	1445	2690.84	.001	.88	.030	.97	.97	1.86
Means Models								
L9 L7_means over time	1452	2586.44	.001	.89	.028	.97	.97	1.78
L10 L7_means gender differences	1452	2870.85	.001	.88	.032	.96	.97	1.98

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; LY = factor loadings; PSI = factor variances and intercorrelations. [a]Metric Model L5 was selected as the best fitting model. [b]Model L7 accepted for substantive hypotheses.

cept the hypothesis of time invariance of factor loadings within gender groups. Model (L3) estimated the factor loadings freely across occasion while testing the hypothesis of gender invariance of factor loadings ($\chi^2 = 2396.18$, $df = 1330$, $p < .001$, GFI = .890, CFI = .974; $\Delta\chi^2 = 67.36$, $df = 34$, $p < .001$). The test of gender invariance was rejected when comparing this model to the configural model. Because time invariance was considered tenable, and gender invariance was not, we evaluated individual gender constraints for all cognitive tests at both time points. The only significant gender differences noted for the 20 cognitive tests were for the Cube Comparison task at Time 1 ($z = -4.88$) and at Time 2 ($z = -3.78$). Therefore, relaxing the equality constraint for the Cube Comparison test between genders, at both time points, provided our test of partial measurement invariance (L4). This model was compared to the configural measurement model and found to have acceptable fit ($\chi^2 = 2373.13$, $df = 1328$, $p < .001$, GFI = .892, CFI = .974; $\Delta\chi^2 = 44.32$, $df = 32$, $p > .07$). With the ability to accept the partially invariant model, we next tested a model adding the longitudinal equivalence of factor loadings. This partially invariant model (L5) tested the hypothesis of simultaneous invariance of factor loadings between genders and across time ($\chi^2 = 2390.97$, $df = 1346$, $p < .001$, GFI = .891, CFI = .974; $\Delta\chi^2 = 62.16$, $df = 50$, $p > .12$) and was tenable. Therefore, the best fitting longitudinal measurement model proved to be L5, with gender and time invariance of all loadings for the cognitive tasks across 7 years with the exception of gender differences in the Cube Comparison task.

Additional tests of substantive hypotheses concerning the factor intercorrelations were conducted. Three models were tested. The first model (L6) added constraints to the accepted partial invariance model to examine time invariance of the factor correlations. The hypothesis of time invariance of factor intercorrelations was rejected ($\chi^2 = 2627.03$, $df = 1388$, $p < .001$, GFI = .879, CFI = .970; $\Delta\chi^2 = 236.06$, $df = 42$, $p < .001$). Model L7 tested gender equivalence of the factor intercorrelations whereas time invariance constraints were removed ($\chi^2 = 2478.83$, $df = 1424$, $p < .001$, GFI = .887, CFI = .974; $\Delta\chi^2 = 87.86$, $df = 78$, $p > .21$), and this model was accepted. A test of gender and time invariance of all factor correlations was also examined and rejected (L8: $\chi^2 = 2690.84$, $df = 1445$, $p < .001$, GFI = .876, CFI = .969; $\Delta\chi^2 = 299.87$, $df = 99$, $p < .001$). In sum, the best fitting measurement model for gender across the adult life span was L5, a partially invariant model. We were also able to accept the hypothesis of gender invariance of the factor intercorrelations, although the hypothesis of equal factor interrelationships across 7 years was rejected (L7). Standardized factor loadings for Model L6 are provided in Table 3. All factor loadings were constrained to be equivalent between genders and across time, with one exception. The cube comparison loading was estimated freely between genders but maintained longitudinal invariance. Loadings are presented separately for genders for cube comparison: women = .500, men = .699. Table 4 contains factor intercorrelations and stabilities, and cross-correlations between the 1984 and 1991 factors are presented in Table 5.

With substantial gender and time equivalence of the factor model established across the adult life span, we next investigated gender differences within three distinct age groups to determine whether differential patterns of gender differences exist in adult cognitive abilities.

Younger Group

The configural model (Y1) freely estimated factor loadings for both gender and occasion ($\chi^2 = 1695.37$, $df = 1296$, $p < .001$, GFI = .778, CFI = .959). The fit of this model was acceptable and more stringent models were examined. The first metric model (Y2) tested time invariance of factor loadings across occasion, also estimating loadings freely for men and women ($\chi^2 = 1735.60$, $df = 1330$, $p < .001$, GFI = .774, CFI = .959; $\Delta\chi^2 = 40.83$, $df = 36$, $p > .26$). The hypothesis of time invariance of factor loadings within gender groups was accepted. A test of gender invariance of factor loadings, Model Y3, allowed the factor loadings to be freely estimated across occasion ($\chi^2 = 1749.71$, $df = 1330$, $p < .001$ GFI = .770, CFI = .957; $\Delta\chi^2 = 55.03$, $df =$

Table 3. Standardized Factor Loadings for the Accepted Partial Invariance Gender Model.

Variable	Inductive Reasoning	Spatial Orientation	Verbal Comprehension	Numeric Facility	Perceptual Speed	Verbal Recall
PMA Reasoning	.852					
ADEPT Letter Series	.791					
Word Series	.811					
Number Series	.660					
PMA Space		.796				
Object Rotation		.832				
Alphanumeric Rotation		.800				
Cube Comparison		.500 – .699[a]				
PMA Verbal Meaning			.455		.563	
ETS Vocabulary			.788			
Advanced Vocabulary			.849			
Word Fluency			.344		.406	
PMA Number				.807		
Addition				.914		
Subtraction and Multiplication				.837		
Identical Pictures					.770	
Number Comparison				.225	.532	
Finding As					.498	
Immediate Recall						.835
Delayed Recall						.838

Note. PMA = Primary Mental Abilities Test; ETS = Educational Testing Services Kit of Factor Referenced Tests; ADEPT = Adult Development and Enrichment Project Test. All factor loadings are significant at $p < .01$.
[a]Cube comparison freely estimated between genders (Women = .500, Men = .699). χ^2 (1424, Women = 540, Men = 442) = 2478.83, GFI = .89, CFI = .97.

36, $p > .02$) and gender invariance of the factor loadings for all six factors was accepted for the young age group. Metric Model Y4, simultaneous invariance of factor loadings between genders and across time, was also accepted (χ^2 = 1764.44, df = 1347, $p < .001$, GFI = .770, CFI = .957; $\Delta\chi^2$ = 69.80, df = 54, $p > .07$). The best fitting longitudinal measurement model (Y4) demonstrated that all factor loadings for the cognitive tasks for our young-old adults were gender and time invariant.

Table 4. Factor Correlations, Stabilities, and Variances for Accepted Longitudinal Gender Model.

	IR	SO	VC	NF	PS	VR	σ_{Time1}	σ_{Time2}
IR	.963	.706	.415	.430	.829	.540	32.86	39.27
SO	.779[a]	.918	.170	.264	.727	.383	117.20[b]	145.41[b]
VC	.451	.235	.978	.229	.184	.344	33.92	34.91
NF	.521[a]	.411[a]	.264	.940	.423	.190	231.73	242.91
PS	.858[a]	.820[a]	.262	.554[a]	.960	.562	37.98[b]	56.12[b]
VR	.648[a]	.538[a]	.407	.323[a]	.683	.791	18.03	21.17

Note. IR = Inductive Reasoning; SO = Spatial Orientation; VC = Verbal Comprehension; NF = Numeric Facility; PS = Perceptual Speed; VR = Verbal Recall; σ = Factor variance. Values above diagonal are 1984. Values below diagonal are 1991. Underlined values indicate factor stabilities.
[a]Factor correlations that changed across time. [b]Factor variances that changed across time.

Table 5. Cross-Correlations for 1984 and 1991 Cognitive Factors.

	IR '91	SO '91	VC '91	NF '91	PS '91	VR '91
IR '84	.....	.727	.475	.514	.779	.588
SO '84	.691	.....	.246	.345	.689	.446
VC '84	.359	.134	.....	.202	.146	.308
NF '84	.383	.260	.239	.....	.374	.181
PS '84	.849	.793	.274	.528	.....	.659
VR '84	.554	.439	.402	.283	.558	.....

Note. IR = Inductive Reasoning; SO = Spatial Orientation; VC = Verbal Comprehension; NF = Numeric Facility; PS = Perceptual Speed; VR = Verbal Recall.

Additional tests of substantive hypotheses concerning the factor intercorrelations were conducted. Three models examining the latent factor space were tested. The first model (Y5) maintained equality constraints of the factor loadings (lambda) between gender groups and over time and the hypothesis of time invariance of factor intercorrelations (psi) was examined ($\chi^2 = 1827.48$, $df = 1389$, $p < .001$, GFI = .760, CFI = .955; $\Delta\chi^2 = 63.02$, $df = 42$, $p > .01$) and accepted. Gender equivalence of the factor intercorrelations with relaxed constraints on time invariance was also found to be plausible (Y6: $\chi^2 = 1849.24$, $df = 1425$, $p < .001$, GFI = .762, CFI = .957; $\Delta\chi^2 = 84.75$, $df = 78$, $p > .28$). Finally, a test of simultaneous gender and time invariance of the factor variance/covariance matrix (Y7: $\chi^2 = 1880.15$, $df = 1446$, $p < .001$, GFI = .755, CFI = .956; $\Delta\chi^2 = 115.65$, $df = 99$, $p > .12$) also proved tenable. Therefore, the substantive hypothesis of gender and longitudinal stability of the relationships between the six cognitive factors was supported. A metric invariance model (Y4) was accepted as the best fitting measurement model for our young adult age group, and support for additional constraints on the relationships between factors was also found (Y7). Summaries of model fit for the young adult group are provided in Table 6.

Middle Age Group

Configural Model M1 estimated factor loadings freely for both gender and occasion ($\chi^2 = 1642.83$, $df = 1296$, $p < .001$, GFI = .800, CFI = .969). Metric Model M2 tested time invariance and freely estimated loadings between genders ($\chi^2 = 1685.98$, $df = 1330$, $p < .001$, GFI = .796, CFI = .969; $\Delta\chi^2 = 48.07$, $df = 36$, $p > .08$). Time invariance of factor loadings was accepted. Gender invariance of factor loadings, Model M3,

Table 6. Comparison of Longitudinal Gender Models for Younger (Y) Age Group.

Model	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	Z-Ratio
Y1 Configural	1296	1695.37	.001	.78	.032	.95	.96	1.31
Y2 Time invariant LY	1330	1735.60	.001	.77	.032	.95	.96	1.30
Y3 Gender invariant LY	1330	1749.71	.001	.77	.033	.95	.96	1.32
Y4 Metric invariant LY	1347	1764.44	.001	.77	.033	.95	.96	1.31
Y5 Time invariant PSI	1389	1827.48	.001	.76	.033	.95	.96	1.32
Y6 Gender invariant PSI	1425	1849.24	.001	.76	.032	.95	.96	1.30
Y7 Metric invariant PSI[a]	1446	1880.15	.001	.76	.032	.95	.96	1.30
Means Models								
Y8 Y7_means over time	1474	1945.44	.001	.76	.033	.95	.95	1.32
Y9 Y7_means between genders	1474	2003.85	.001	.78	.035	.94	.95	1.36

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; LY = factor loadings; PSI = factor variances and intercorrelations.
[a]Metric Model Y7 was selected as the best fitting model.

freely estimated factor loadings across occasions (χ^2 = 1683.99, *df* = 1330, *p* < .001, GFI = .796, CFI = .969; $\Delta\chi^2$ = 48.10, *df* = 36, *p* > .08) and gender invariance of the factor loadings was accepted for the middle age group. Model M4 tested gender and time invariance of factor loadings (χ^2 = 1708.85, *df* = 1347, *p* < .001, GFI = .793, CFI = .968; $\Delta\chi^2$ = 73.09, *df* = 54, *p* > .04). This model demonstrated gender and time invariant factor loadings for all cognitive tasks and was accepted as the best fitting measurement model for the middle age group.

Model M5 tested equivalence of factor correlations across time, and values were freely estimated between genders (χ^2 = 1765.39, *df* = 1389, *p* < .001, GFI = .788, CFI = .967; $\Delta\chi^2$ = 56.91, *df* = 42, *p* > .06). The test of gender equivalence of the factor space relaxed the time invariance constraint (M6: χ^2 = 1803.19, *df* = 1425, *p* < .001, GFI = .778, CFI = .967; $\Delta\chi^2$ = 94.57, *df* = 78, *p* > .09) and was accepted. Finally, both gender and time equivalence of the factor correlation matrix was examined (M7: χ^2 = 1826.48, *df* = 1446, *p* < .001, GFI = .777, CFI = .964; $\Delta\chi^2$ = 118.06, *df* = 99, *p* > .09). As with the younger age group, the substantive hypothesis of gender and longitudinal stability of the relationships between the six cognitive factors was supported in the middle age group. Metric invariance Model M4 was accepted as the best fitting measurement model for our 50 – 63 year-old adults. Furthermore, additional gender and time constraints on the factor space were also supported. Therefore, metric Model M7 was accepted as best fitting for the middle age group. Information about model fit for the middle age group adults is found in Table 7.

Older Group

Configural Model O1 fit well for the older age group (χ^2 = 1715.39, *df* = 1296, *p* < .001, GFI = .804, CFI = .969). Metric Model O2, time invariance of factor loadings, was also accepted (χ^2 = 1756.55, *df* = 1330, *p* < .001, GFI = .799, CFI = .969; $\Delta\chi^2$ = 41.90, *df* = 36, *p* > .23). Gender invariance of factor loadings, tested in Model O3, also provided nonsignificant decrement in model fit (χ^2 = 1758.29, *df* = 1330, *p* < .001, GFI = .798, CFI = .969; $\Delta\chi^2$ = 42.79, *df* = 36, *p* > .20). Finally, Model O4, gender and time invariance of factor loadings, was also accepted (χ^2 = 1786.28, *df* = 1347, *p* < .001, GFI = .795, CFI = .968; $\Delta\chi^2$ = 79.97, *df* = 54, *p* > .06). The gender and time invariant Model O4 was accepted as the best fitting measurement model for the older participants.

Model O5 tested time invariance of factor correlations, and values for men and women were freely estimated (χ^2 = 1872.04, *df* = 1389, *p* < .001, GFI = .787, CFI = .965; $\Delta\chi^2$ = 85.45, *df* = 42, *p* < .001). A lack of time invariance in the factor intercorrelations for our oldest age group was noted. However, gender equivalence Model O6 of the factor space was accepted (χ^2 = 1874.71, *df* = 1425, *p* < .001, GFI = .789, CFI = .967; $\Delta\chi^2$ = 87.93, *df* = 78, *p* > .20). Finally, the hypothesis of gender and time equivalence of the factor correlation matrix was also rejected

Table 7. Comparison of Longitudinal Gender Models for Middle (M) Age Group.

Model		*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	Z-Ratio
M1	Configural	1296	1642.83	.001	.80	.028	.96	.97	1.27
M2	Time invariant LY	1330	1685.98	.001	.80	.029	.96	.97	1.27
M3	Gender invariant LY	1330	1683.99	.001	.80	.029	.96	.97	1.27
M4	Metric invariant LY	1347	1708.85	.001	.79	.029	.96	.97	1.27
M5	Time invariant PSI	1389	1765.39	.001	.79	.029	.96	.97	1.27
M6	Gender invariant PSI	1425	1803.19	.001	.78	.028	.96	.97	1.27
M7	Metric invariant PSI[a]	1446	1826.48	.001	.78	.028	.96	.97	1.26
Means Models									
M8	M7_means over time	1474	1915.42	.001	.77	.030	.96	.96	1.30
M9	M7_means between genders	1474	2028.66	.001	.77	.034	.95	.95	1.38

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; LY = factor loadings; PSI = factor variances and intercorrelations.
[a]Metric Model M7 was selected as the best fitting model.

(O7: χ^2 = 1939.92, *df* = 1446, $p < .001$, GFI = .782, CFI = .964; $\Delta\chi^2$ = 153.17, *df* = 99, $p < .001$). Differing from the young and middle age groups, gender but not time invariance of the factor intercorrelations was accepted for our 64 – 87 year-old adults. As in younger groups, metric invariance Model O4 was accepted as the best fitting measurement model; however, differences were noted in the longitudinal factor space for the oldest group. Therefore, metric Model O6, demonstrating gender invariance and a lack of time invariance between latent factors, was accepted as best fitting for the oldest age group. A summary of model fit for the older age group is provided in Table 8. Standardized factor loadings are presented for all three age groups in Table 9. Table 10 provides standardized factor intercorrelations for all three age groups. The younger and middle age groups were gender and time invariant, whereas the older participants' factor correlations only demonstrated gender invariance. Note that 14 of 15 factor intercorrelations were larger at the second time point for the older group; however, only 3 statistically significant increases, all involving Perceptual Speed, were found (Spatial Orientation – Perceptual Speed, Numerical Facility – Perceptual Speed, Perceptual Speed – Verbal Recall). Factor variances and stabilities for all age group models are provided in Table 11. Whereas factor variances were stable across time in the younger and middle age groups, increased factor variances were noted for the oldest participants; however, variability of perceptual speed was the only statistically significant increase.

Latent Means Model Extensions

Models accepted as best fitting for the covariance structures have been described above for longitudinal and multiple-group analyses. Whereas these models demonstrate the equivalence of measurement properties of the factor model, these analyses provide no information about gender differences and longitudinal change for the six cognitive abilities. Therefore, means from the observed measures were added to the accepted covariance structures to estimate latent mean gender differences and change over 7 years. For the alpha matrices *t* values provided by LISREL were used to determine statistically significant differences. Significance level for all latent factor mean comparisons was ± 2.58, $p < .01$.

Between Gender Latent Mean Differences

The accepted longitudinal measurement model from the overall gender comparison (partial invariant Model L7) was extended to estimate latent factor means. Between group variation was tested by setting alpha values for all estimated cognitive ability factors to zero for women. Scores for men are estimated as deviation values, using women as a reference group. It should be noted that intercepts are freely estimated for each time point but are constrained to be equivalent between gender groups. Table 12 provides the mean deviation values (alphas), standard

Table 8. Comparison of Longitudinal Gender Models for Older (O) Age Group.

Model		*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	Z-Ratio
O1	Configural	1296	1715.40	.001	.80	.030	.96	.97	1.32
O2	Time invariant LY	1330	1756.55	.001	.80	.030	.96	.97	1.32
O3	Gender invariant LY	1330	1758.29	.001	.80	.030	.96	.97	1.32
O4	Metric invariant LY	1347	1786.28	.001	.80	.030	.96	.97	1.33
O5	Time invariant PSI	1389	1872.04	.001	.79	.031	.96	.97	1.35
O6	Gender invariant PSI[a]	1425	1874.71	.001	.79	.030	.96	.97	1.32
O7	Metric Invariant PSI	1446	1939.92	.001	.78	.031	.96	.96	1.34
Means Models									
O8	O6_means over time	1453	1948.04	.001	.79	.031	.96	.96	1.34
O9	O6_means between genders	1453	1993.97	.001	.79	.032	.96	.96	1.37

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; LY = factor loadings; PSI = factor variances and intercorrelations.
[a]Metric Model O6 was selected as the best fitting model.

Table 9. Standardized Solutions for Age by Gender Analyses.

Factors/Variables	Younger	Middle	Older
Inductive Reasoning			
PMA Reasoning	.771	.814	.823
ADEPT Letter Series	.731	.753	.733
Word Series	.736	.722	.789
Number Series	.543	.639	.630
Spatial Orientation			
PMA Space	.785	.755	.783
Object Rotation	.768	.786	.830
Alphanumeric Rotation	.665	.732	.777
Cube Comparison	.517	.350	.458
Verbal Comprehension			
PMA Verbal Meaning	.609	.509	.365
ETS Vocabulary	.796	.757	.797
Advanced Vocabulary	.832	.344	.863
Word Fluency	.359	.364	.350
Numerical Facility			
PMA Number	.786	.810	.820
Addition	.926	.919	.900
Subtraction + Multiplication	.826	.806	.867
Number Comparison	.157	.171	.256
Perceptual Speed			
Identical Pictures	.507	.565	.686
Number Comparison	.506	.460	.464
Finding As	.552	.494	.565
Word Fluency	.332	.288	.318
PMA Verbal Meaning	.225	.453	.589
Verbal Recall			
Immediate Recall	.797	.828	.832
Delayed Recall	.798	.796	.810

errors, and information concerning statistically significant gender differences for the six cognitive abilities for 1984 and 1991 respectively.

Significant gender differences in latent factor means were noted for Spatial Orientation and Verbal Recall factors at both time points. Women did significantly worse than men on Spatial Ability and outperformed men on Verbal Recall. The pattern for 1991 was similar although the difference in Spatial Ability diminished and a slightly larger deviation was noted between genders for Verbal Recall. Figure 1 displays similar deviation patterns for both time points.

Longitudinal Change in Latent Means

Changes in latent factor means were calculated by fixing the values for all 1984 latent means to zero to serve as a reference point. This model estimates values for 1991 latent means as deviation scores between 1984 and 1991. For the factor change scores *t* values are plotted and represent ability-specific mean-level change between 1984 and 1991. Figure 2 illustrates these deviation values. Estimated *t* values for within-group change (Table 13) show that both men and women experienced significant latent mean declines for the Spatial Orientation, Inductive Reasoning, Numerical Facility, and Perceptual Speed constructs. Verbal Comprehension and Verbal Recall performance remained stable for both men and women.

Gender Differences in Latent Means by Age Group

A consistent pattern of gender differences was

Table 10. Rescaled Factor Intercorrelations for Age by Gender Analyses.

Factors	Younger (Invariant)	Middle (Invariant)	Older 1984	Older 1991
Inductive Reasoning/Spatial Orientation	.554	.587	.597	.710
Inductive Reasoning/Verbal Comprehension	.368	.543	.636	.634
Inductive Reasoning/Numerical Facility	.472	.442	.528	.593
Inductive Reasoning/Perceptual Speed	.715	.693	.772	.841
Inductive Reasoning/Verbal Recall	.308	.382	.532	.621
Spatial Orientation/Verbal Comprehension	.057	.259	.299	.380
Spatial Orientation/Numerical Facility	.255	.304	.301	.462
Spatial Orientation/Perceptual Speed	.515	.555	.671[a]	.793[a]
Spatial Orientation/Verbal Recall	.216	.119	.355	.490
Verbal Comprehension/Numerical Facility	.149	.176	.337	.385
Verbal Comprehension/Perceptual Speed	.198	.353	.364	.459
Verbal Comprehension/Verbal Recall	.344	.414	.435	.523
Numerical Facility/Perceptual Speed	.583	.551	.619[a]	.740[a]
Numerical Facility/Verbal Recall	.095	.135	.319	.444
Perceptual Speed/Verbal Recall	.474	.336	.492[a]	.631[a]

[a]Indicates correlations that differed between time points.

found across the three age groups. Men performed at a higher level for Spatial Orientation in all three age groups, and at both time points. Women outperformed men on Verbal Recall for all age groups and at both occasions. Women in the younger group performed significantly better than younger men in Perceptual Speed in 1991 only. Middle-aged women performed significantly better than middle-aged men in Perceptual Speed for both 1984 and 1991. In contrast, gender differences for Perceptual Speed were absent in the oldest group (see Table 12).

Longitudinal Change in Latent Means by Age Group

Younger group

Models assessing 7-year changes in latent factor means are summarized in Table 13. Overall trends for the younger group showed a pattern of either stability or improvement in performance over time. Younger women improved on Inductive Reasoning and Verbal Comprehension whereas younger men did better in Spatial Orientation and Verbal Comprehension. No statistically significant decline on any latent factor was noted for either gender in the youngest group.

Table 11. Factor Variances and Stabilities for Cohort by Gender Analyses.

Factors	Younger		Middle		Older		
	σ	Stability	σ	Stability	σ_{Time1}	σ_{Time2}	Stability
Inductive Reasoning	18.38	.967	24.35	.967	25.06	27.76	.932
Spatial Orientation	79.96	.954	86.88	.881	111.65	123.74	.859
Verbal Comprehension	31.24	.995	34.10	.998	35.98	39.35	.970
Numerical Facility	235.36	.976	203.71	.955	220.30	251.86	.929
Perceptual Speed	11.46	.980	14.88	.893	18.94	28.13[a]	.899
Verbal Recall	13.35	.721	14.25	.739	13.89	16.65	.731

Note. σ = factor variance.
[a]Statistically significant increase in factor variance.

Table 12. Mean Deviation Values, Standard Errors, and t Values for Gender Differences Models.

Group	Inductive Reasoning	Spatial Orientation	Verbal Comprehension	Numerical Facility	Perceptual Speed	Verbal Recall
All Subjects 1984						
Alpha	.030	5.338[a]	.571	1.378	−1.000	−1.454[b]
Standard Error	(.381)	(.729)	(.395)	(1.016)	(.441)	(.283)
t value	.079	7.323	1.444	1.357	−2.267	−5.136
All Subjects 1991						
Alpha	−.134	5.248[a]	.315	.732	−1.043	−1.610[b]
Standard Error	(.412)	(.799)	(.398)	(1.033)	(.511)	(.306)
t value	−.362	6.570	.791	.708	−2.042	−5.254
Younger 1984						
Alpha	.161	4.492[a]	.187	.634	−1.156	−1.348[b]
Standard Error	(.538)	(1.147)	(.694)	(1.862)	(.471)	(.449)
t value	.298	3.918	.269	.341	−2.455	−3.000
Younger 1991						
Alpha	−.162	4.869[a]	.179	.412	−1.245[b]	−1.354[b]
Standard Error	(.530)	(1.140)	(.682)	(1.842)	(.460)	(.456)
t value	−.307	4.273	.263	.224	−2.703	−2.971
Middle 1984						
Alpha	−.013	4.780[a]	1.149	2.084	−1.395[b]	−1.166[b]
Standard Error	(.573)	(1.133)	(.681)	(1.648)	(.500)	(.446)
t value	−.022	4.220	1.687	1.265	−2.787	−2.613
Middle 1991						
Alpha	.141	5.562[a]	.804	2.008	−1.343[b]	−1.816[b]
Standard Error	(.570)	(1.130)	(.682)	(1.640)	(.495)	(.451)
t value	.247	4.921	1.178	1.224	−2.713	−4.026
Older 1984						
Alpha	−.720	5.292[a]	.281	.802	−1.225	−2.003[b]
Standard Error	(.566)	(1.222)	(.675)	(1.658)	(.532)	(.423)
t value	−1.272	4.331	.417	.483	−2.302	−4.730
Older 1991						
Alpha	−1.153	3.749[a]	−.254	−1.316	−1.349	−1.987[b]
Standard Error	(.593)	(1.267)	(.701)	(1.756)	(.618)	(.459)
t value	−1.944	2.959	−.362	−.749	−2.182	−4.328

Note. t values less than ± 2.58 are not statistically significant.
[a]Positive significant values indicate gender differences favoring men. [b]Negative values demonstrate women performing better than men.

Middle group

Men's cognitive abilities remained stable except for statistically significant declines in Numerical Facility and Perceptual Speed. In contrast, only Verbal Recall remained stable over 7 years for women, and declines in Inductive Reasoning, Spatial Orientation, Numerical Facility, and Perceptual Speed were noted. Conversely, middle-aged women showed significant improvement on Verbal Comprehension.

Older group

Older participants declined on all latent cognitive factors with one exception, Verbal Comprehension for older women. In summary, clear patterns emerged showing both gender differences and longitudinal change in latent factor means. Figure 3 illustrates longitudinal mean change for the overall life span, demonstrating consistency of the gender by age group change patterns.

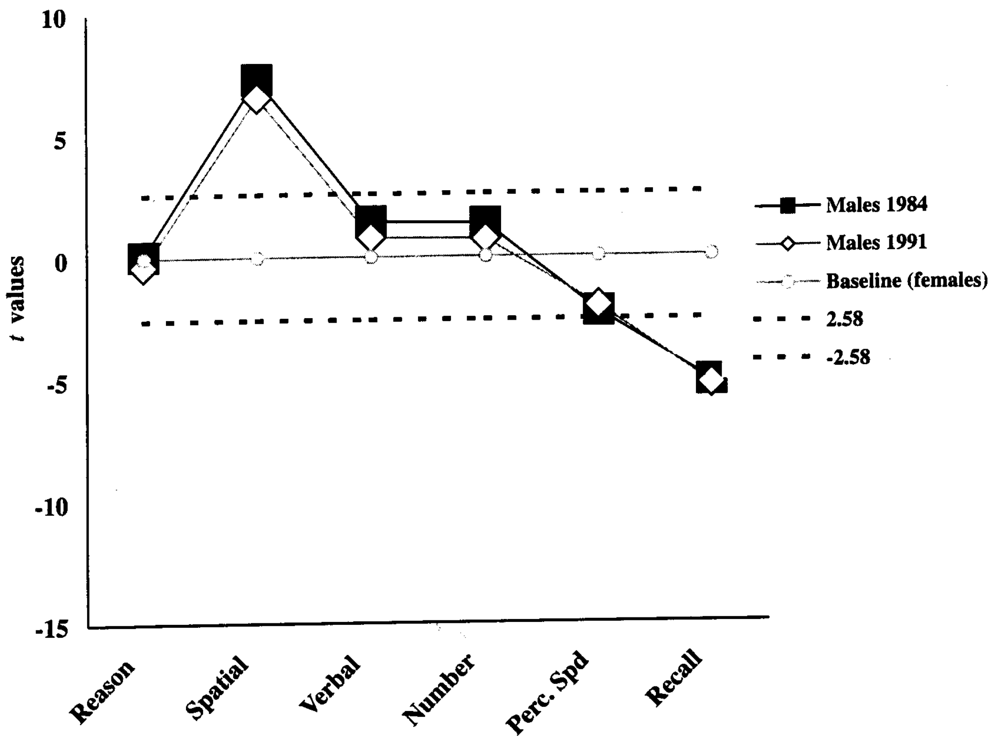

Fig. 1. Gender differences for latent means for six cognitive factors. Latent means for women in 1984 and 1991 were set to zero and used as a reference group to derive deviation values for male latent means for 1984 and 1991 respectively. Boundaries indicate level of significance for $p < .01$. Perc. Spd = Perceptual Speed.

DISCUSSION

Gender differences in psychometric abilities have long been discussed in the literature. Many of the earlier studies used observed scores to test for gender differences, often doing so without regard for issues of measurement equivalence. Information concerning the latent structure of cognitive abilities for men and women across the middle and later years of the life span is relatively sparse (e.g., Schaie et al., 1998). The aim of the current study was to examine the structure and dimensions of a six-factor model of cognitive abilities between gender groups across the adult life span and to provide detailed information about age-group specific changes in cognitive abilities.

Results illustrate the importance of the issue of measurement invariance when assessing gender differences in cognitive abilities. Demonstrating equivalence of factor loadings between groups or across occasions is relevant for all aspects of developmental research; however, demonstrations of invariance of measurement structures are rare. Invariance of constructs has been described as the logical opposite of change (Cunningham, 1991). Without evidence of the metric equivalence of constructs, quantitative inferences about properties of variables or scales may not be warranted. Measurement invariance ensures that comparable units of measurement exist and allows for meaningful quantitative comparisons of groups or across time (Cunningham, 1982, 1991; Horn, 1991; Horn & McArdle, 1992; Horn et al., 1983).

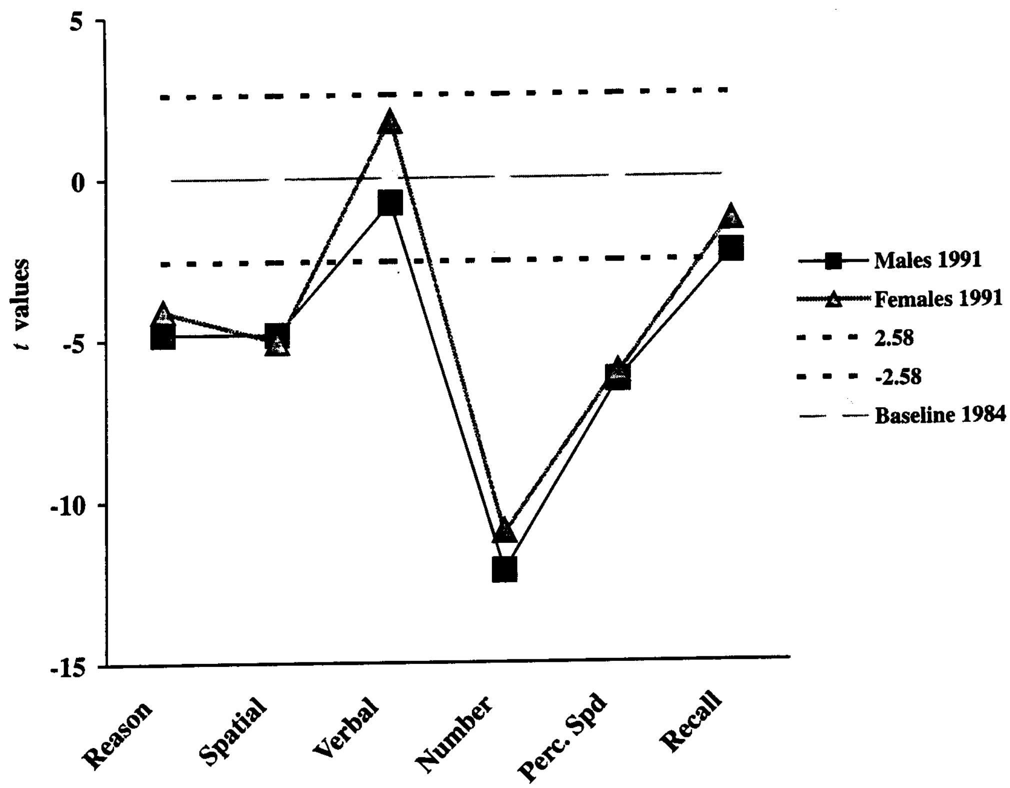

Fig. 2. Longitudinal change in latent means for six cognitive factors from 1984 to 1991. No significant change in latent means was noted for men and women on the Verbal Comprehension and Verbal Recall factors. Boundaries indicate level of significance for $p < .01$. Perc. Spd = Perceptual Speed.

Schaie et al. (1998) demonstrated measurement invariance was plausible in cognitive abilities across the majority of the adult life span. Solutions for the youngest (aged 22 – 42) and oldest (aged 71 – 87) participants in that study were estimated separately from the remaining four cohort groups (aged 43 – 70). Direct tests of gender differences or gender-specific changes across time in the factor model were not conducted. The current study provides one of the few examinations of gender differences and longitudinal changes in latent cognitive abilities for the entire adult life span. Furthermore, use of structural equation modeling techniques allowed a direct test of measurement invariance of these cognitive abilities, providing assurance that the factors are directly comparable between genders and across time.

Results from the entire sample provide evidence for gender and time invariance across 7 years. The Cube Comparison test proved the only cognitive measure demonstrating significant differential measurement properties between genders. Allowing this test to estimate freely between genders resulted in a partial invariance measurement model being accepted as best fitting the cognitive abilities across the adult life span. Hertzog and Carter (1982) demonstrated similar gender differences using Raven's matrices, and offered an explanation that women are more likely to use a nonspatial strategy for making inferences about hidden cube faces. However, this condition of partial invariance only existed when the entire adult life span was examined simultaneously. Separate analyses within our three age groups indicated

Table 13. Mean Deviation Values, Standard Errors, and *t* Values for Longitudinal Change Models.

Group	Inductive Reasoning	Spatial Orientation	Verbal Comprehension	Numerical Facility	Perceptual Speed	Verbal Recall
All Males						
Alpha	–.598	–1.467	–.085	–3.925	–1.151	–.358
Standard Error	(.124)	(.303)	(.112)	(.323)	(.184)	(.155)
t value	–4.819	–4.849	–.757	–12.153	–6.264	–2.314
All Females						
Alpha	–.460	–1.451	.181	–3.270	–1.035	–.181
Standard Error	(.112)	(.286)	(.103)	(.299)	(.172)	(.138)
t value	–4.103	–5.080	1.759	–10.942	–6.003	–1.313
Young Males						
Alpha	.295	1.363	.761	–.669	.122	.581
Standard Error	(.190)	(.410)	(.176)	(.476)	(.211)	(.274)
t value	1.557	3.322	4.331	–1.406	.577	2.117
Young Females						
Alpha	.576	.997	.761	–.574	.195	.574
Standard Error	(.179)	(.406)	(.153)	(.426)	(.204)	(.237)
t value	3.223	2.459	4.969	–1.346	.957	2.421
Middle Males						
Alpha	–.370	–1.014	.100	–2.966	–1.039	–.402
Standard Error	(.190)	(.510)	(.161)	(.474)	(.254)	(.252)
t value	–1.949	–1.986	.622	–6.263	–4.093	–1.593
Middle Females						
Alpha	–.511	–1.804	.471	–2.935	–.922	.266
Standard Error	(.178)	(.482)	(.173)	(.462)	(.245)	(.232)
t value	–2.869	–3.743	2.724	–6.353	–3.768	1.146
Older Males						
Alpha	–1.655	–4.718	–.957	–7.524	–1.974	–1.052
Standard Error	(.223)	(.601)	(.215)	(.577)	(.284)	(.252)
t value	–7.433	–7.854	–4.451	–13.048	–6.934	–4.179
Older Females						
Alpha	–1.243	–3.235	–.441	–5.518	–1.788	–1.080
Standard Error	(.193)	(.532)	(.183)	(.528)	(.254)	(.219)
t value	–6.436	–6.082	–2.412	–10.449	–7.049	–4.943

Note. t values less than ± 2.58 are not statistically significant and indicate mean stability of the latent construct over a 7-year period. Positive values indicate better performance, and negative values indicate decline.

complete metric invariance of all factor loadings, including the Cube Comparison test. Increased variability when examining gender differences across the entire adult life span is the likely explanation for this result. With the noted exception of the Cube Comparison task in the overall gender comparison, remarkable stability of the covariance structure was demonstrated, providing confidence that these cognitive abilities may be compared between genders.

Testing for gender differences within three age groups demonstrated differential patterns of gender differences and short-term longitudinal change. Metric invariance of factor loadings was accepted for all three age groups. Additionally, younger and middle-aged adult groups had equivalent factor variances and covariances between genders *and* over a 7-year period. For the oldest group, participants aged 64 to 87, factor variances and intercorrelations were equivalent across gender groups whereas these values between latent factors had to be estimated separately for each time point. The demonstration of metric invariance of the factor loadings was im-

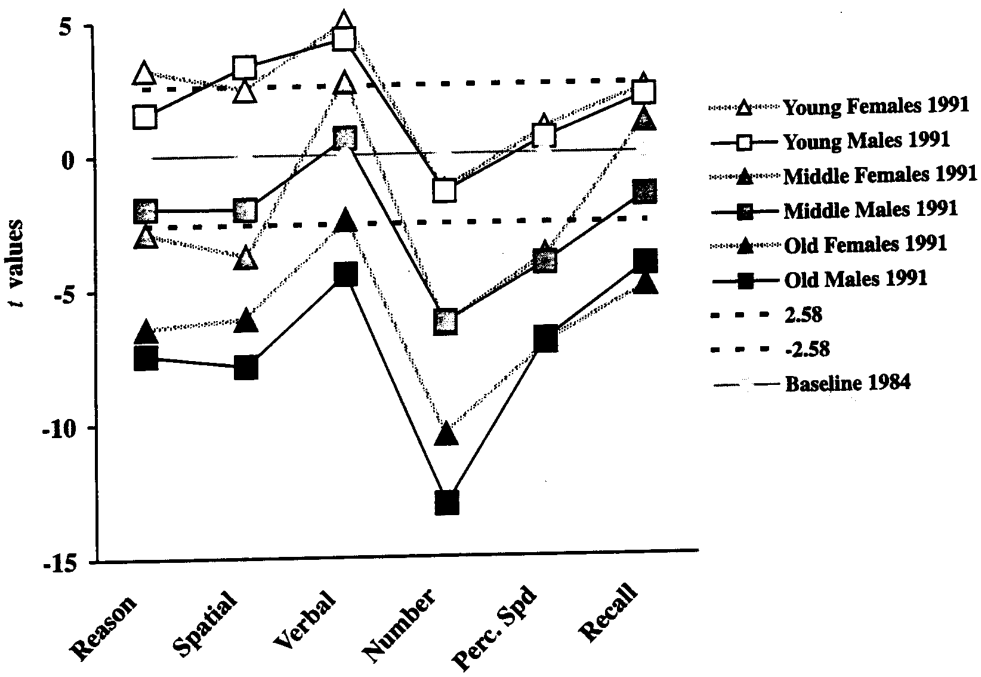

Fig. 3. Longitudinal latent mean change for gender by age group models. Boundaries indicate level of significance for $p < .01$. Perc. Spd = Perceptual Speed.

portant because it permits group and time comparisons.

Relationships between the latent constructs for the oldest participants showed increased variability over 7 years. Ninety-three percent (14 of 15) of the factor intercorrelations were larger at Time 2 than Time 1; however, only 3 or 21% of those increases were statistically significant. All significant increases involved relationships with the Perceptual Speed factor. Furthermore, a significant increase in the factor variance for perceptual speed was also noted for the oldest participants. In summary, results from the age-group specific covariance analyses illustrated substantial gender and time invariance. However, an important caveat is noted. The lack of gender invariance of factor loadings across the entire adult life span was corrected by implementing partial invariance techniques. Surprisingly, only one indicator (at two time points) or 5% of the factor loadings differed between genders. Developmental researchers should be aware that a single test in a battery may create enough leverage to cause a tenable model to be rejected. Examining unstandardized factor loadings for statistically significant differences, and adjusting for those differences, may allow acceptance of a more stringent level of invariance.

These analyses are an interesting example of the process of cognitive development across the adult life span. Controlling for age or more specifically, examining adults in age groups that span later life, demonstrated that men and women do not show differential measurement structures of their cognitive abilities. These findings provide strong support for the similarity of the structure of cognitive abilities between genders and are consistent with earlier work (e.g., Hertzog & Carter, 1982). As noted by Schaie et al. (1998), increased variability in the oldest and youngest participants may lead to a lack of metric invariance in overall-sample models. However, separate age group analyses in the current study demonstrated invariance of factor loadings and gender invariance of the factor variances and covariances. Larger factor correlations and increasing factor variances over time in the oldest age group may provide additional support for the hypothesis of dedifferentiation of cognitive abilities, but only in adults over age 70. Localization of the specific lack of fit, or reason for rejection of tests of invariance, is critical and highly informative.

Consistent gender differences in the latent means support previously published reports of observed gender differences in cognitive abilities at various ages. Women performed significantly worse than men for Spatial Ability, and generally outperformed men on Verbal Recall and in some instances in Perceptual Speed. Gender differences in Perceptual Speed were not present when examined across the entire adult life span. Age-specific gender differences (favoring women) in this ability were found for only the young and middle age groups. Levels of performance for Numeric Facility, Inductive Reasoning, and Verbal Comprehension were not significantly different between men and women. The equivalence of the covariance structures and the general pattern of nonsignificant mean differences between genders indicate a high degree of similarity in adult cognitive abilities. Remarkable stability of longitudinal between-gender differences was also noted for the entire sample.

How do these results compare to previous studies? These data are not completely consistent with those presented by Schaie and Hertzog (1983), who observed differences on PMA spatial and number abilities favoring men, and word fluency favoring women. However, Schaie and Hertzog used only single-marker tests of each ability whereas the current study examined latent constructs. Multiply marked latent factors reduce the potential measurement error that is inherent in single-marker tests and may partially explain the differences between these studies. The current study found no gender differences in Numerical Facility; however, our results replicate the spatial ability findings of Schaie and Hertzog. Furthermore, women performed better than men on Perceptual Speed. The word fluency test examined by Schaie and Hertzog loads on the Perceptual Speed factor in the current study, indirectly supporting the gender difference for this task.

Our findings for gender differences in spatial ability support those of previous studies (Cohen, Schaie, & Gribbin, 1977; Linn & Petersen, 1985, 1986). Furthermore, most of our spatial ability indicators are tests of mental rotation, which normally produce larger gender differences than do spatial perception or spatial visualization tasks (Hyde & McKinley, 1997; Linn & Peterson, 1985). These data also support the meta-analysis of Hyde and Linn (1988), which reported a lack of gender differences in verbal ability. Additionally, our gender differences for Perceptual Speed are similar to those reported by Feingold (1988) and parallel findings obtained for the perceptual speed subtests (Digit Symbol) of the Wechsler scales (Feingold, 1993). Similar findings have been documented by Born, Bleichrodt, and Van Der Flier (1987) and are consistent with the classic reviews that document the female superiority in Perceptual Speed (Anastasi, 1958; Maccoby & Jacklin, 1974; Tyler, 1965). However, the fine grained analysis of three age groups revealed an attenuation of differences in Perceptual Speed in the oldest age group. Whereas women have an advantage in Perceptual Speed throughout most of adulthood, they lose this advantage after age 70. The observed difference favoring women on Verbal Recall contradicts (Born et al., 1987; Feingold, 1993; Maccoby & Jacklin, 1974) and supports (Herlitz et al., 1998; Hultsch et al., 1998; Schaie & Willis, 1993; Zelinski et al., 1993) previous findings.

Declines on 4 of the 6 cognitive factors were noted for both genders. Verbal Comprehension and Verbal Recall were the only factors to display longitudinal mean stability for both men and women. Similarity in the patterns of change over time in latent cognitive factors for the total sample was remarkable. Does this suggest that observed decline in latent cognitive abilities is occurring universally across the adult life span? The answer to this question can be found by examining longitudinal change in latent factor means by age group. For the younger group, both men and women remained stable or improved their performance over the 7-year period. Men in the middle age group were stable on 4 cognitive abilities in contrast to women who declined on 4 abilities. However, women showed statistically significant improvement in Verbal Comprehension. Finally, the older age group demonstrated statistically significant decline across all ability factors, with stability on Verbal Comprehension for older women the only exception.

Kaufman et al. (1991) suggested that there has been little systematic attention to whether men and women show similar patterns of intellectual change over the life span. Summarizing problems with research on gender differences in cognitive abilities, Caplan and Caplan (1997) suggested that differences are not systematically or invariably found and do not exist at all ages. Furthermore, Hertzog and Carter (1982) suggested their results warranted testing invariance of cognitive structures in a test battery explicitly designed to estimate primary ability factors. Evidence presented here provides both systematic and detailed information concerning the invariant structure of primary ability factors, and patterns of intellectual change and gender differences across the adult life span. The current study demonstrated:

1. The latent structure of cognitive abilities is the same for men and women across the adult life span.
2. Latent cognitive abilities can be reliably measured longitudinally.
3. Ignoring age-group specific analyses, over a 7-year period men and women declined significantly in 4 of 6 latent cognitive abilities. However, younger men and women remained stable or improved, middle age group participants remained stable or declined, and older participants generally performed worse 7 years later.
4. Women outperformed men in Perceptual Speed and Verbal Recall whereas men were significantly better in Spatial Orientation.
5. No gender differences in latent means were noted for Inductive Reasoning, Numeric Facility, and Verbal Comprehension across the adult life span ($\text{ages}_{\text{Time1}}$ 22 – 87; $\text{ages}_{\text{Time2}}$ 29 – 94).

What do these results concerning measurement invariance and latent means suggest for future research? First, these findings demonstrate measurement invariance of cognitive abilities for men and women across the adult life span. Second, testing measurement invariance within a covariance structure provides only a partial understanding of gender differences of cognitive abilities. Analysis of both the covariance and mean structures permits a thorough examination of both the measurement structure and the level of performance differences. Modeling latent factor means allows inferences about gender differences and longitudinal change to be made about specific latent abilities. The current study addressed differences in patterns of change over 7 years by examining the entire sample simultaneously and separately in three distinct age groups. The shape of the change profiles was remarkably similar for genders and across all age groups; however, differences in mean levels are clearly demonstrated. In conclusion, substantial evidence for gender invariance of cognitive abilities was demonstrated across the adult life span.

REFERENCES

Alwin, D.F. (1988). Structural equation models in research on human development and aging. In K.W. Schaie, R.T. Campbell, W. Meredith, & S.C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 71-170). New York: Springer Publishing.

Anastasi, A. (1958). *Differential psychology* (3rd ed.). New York: MacMillan.

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.

Blieszner, R., Willis, S.L., & Baltes, P.B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2*, 247-265.

Bollen, K.A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Born, M.P., Bleichrodt, N., & Van Der Flier, H. (1987). Cross-cultural comparison of sex-related differences on intelligence tests: A meta-analysis. *Journal of Cross Cultural Psychology, 18,* 283-314.

Browne, M.W., & Cudek, R. (1993). Alternate ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B.M. (1998). Application 8: Testing for invariant latent mean structures. In *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming* (pp. 303-325). Mahwah, NJ: Lawrence Erlbaum.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456-466.

Caplan, P.J., & Caplan, J.B. (1997). Do sex-related cognitive differences exist, and why do people seek them out? In P.J. Caplan, M. Crawford, J.S. Hyde, & J.T.E. Richardson (Eds.), *Gender differences in human cognition* (pp. 52-80). New York: Oxford University Press.

Carmines, E.G., & McIver, J.P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G.W. Bohrnnstedt & E.F. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65-115). Beverly Hills, CA: Sage.

Cohen, D., Schaie, K.W., & Gribbin, K. (1977). The organization of spatial ability in older men and women. *Journal of Gerontology, 32,* 578-585.

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* (pp. 106-113). Washington, DC: American Psychological Association.

Deaux, K. (1985). Sex and gender. *Annual Review of Psychology, 36,* 49-81.

Ekstrom, R.B., French, J.W., Harman, H., & Derman, D. (1976). *Kit of factor referenced cognitive tests* (rev ed.). Princeton, NJ: Educational Testing Service.

Everson, H.T., Millsap, R.E., & Rodriguez, C.M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the test anxiety inventory. *Educational and Psychological Measurement, 51,* 243-251.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95-103.

Feingold, A. (1993). Cognitive gender differences: A developmental perspective. *Sex Roles, 29,* 91-112.

Halpern, D.F. (1992). *Sex differences in cognitive abilities.* Hillsdale, NJ: Lawrence Erlbaum.

Hart, R.P., & O'Shanick, G.J. (1993). Forgetting rates for verbal, pictorial, and figural stimuli. *Journal of Clinical and Experimental Neuropsychology, 15,* 245-265.

Herlitz, A., Nilsson, L.G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory and Cognition, 25,* 801-811.

Hertzog, C., & Carter, L. (1982). Sex differences in the structure of intelligence: A confirmatory factor analysis. *Intelligence, 6,* 287-303.

Horn, J.L. (1991). Comments on issues in factorial invariance. In L.M. Collins & J.L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114-125). Washington, DC: American Psychological Association.

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.

Hultsch, D.F., Hertzog, C., Dixon, R.A., & Small, B.J. (1998). *Memory change in the aged.* New York: Cambridge University Press.

Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53-69.

Hyde, J.S., & McKinley, N.M. (1997). Gender differences in cognition: Results from meta-analyses. In P.J. Caplan, M. Crawford, J.S. Hyde, & J.T.E. Richardson (Eds.), *Gender differences in human cognition* (pp. 30-51). New York: Oxford University Press.

James, L.R., Muliak, S.A., & Brett, J.M. (1982). *Causal analysis: Assumptions, models and data.* Beverly Hills, CA: Sage.

Jöreskog, K.G. (1979). Statistical estimation of structural models in longitudinal developmental investigations. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303-351). New York: Academic Press.

Jöreskog, K.G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt Books.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications.* Chicago: SPSS.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8* (Version 8.20) [Computer software]. Chicago: Scientific Software International.

Kaufman, A.S., Kaufman-Packer, J.L., McLean, J.E., & Reynolds, C.R. (1991). Is the pattern of intellectual growth and decline across the adult life-span different for men and women? *Journal of Clinical Psychology, 47,* 801-812.

Linn, M.C., & Petersen, A.C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479-1498.

Linn, M.C., & Petersen, A.C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J.S. Hyde & M.C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore, MD: Johns Hopkins University Press.

Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Meinz, E.J., & Salthouse, T.A. (1998). Is age kinder to females than to males? *Psychonomic Bulletin and Review, 5,* 56-70.

Millsap, R., & Everson, H. (1991). Confirmatory measurement model comparison using latent means. *Multivariate Behavioral Research, 26,* 479-497.

Rosén, M. (1995). Gender differences in structure, means and variances of hierarchically ordered ability dimensions. *Learning and Instruction, 5,* 37-62.

Schaie, K.W. (1985). *Manual for the Schaie-Thurstone Adult Mental Abilities Test (STAMAT).* Palo Alto, CA: Consulting Psychologists Press.

Schaie, K.W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.

Schaie, K.W., Dutta, R., & Willis, S.L. (1991). The relationship between rigidity-flexibility and cognitive abilities in adulthood. *Psychology and Aging, 6,* 371-383.

Schaie, K.W., & Hertzog, C. (1983). Fourteen-year cohort-sequential analyses of adult intellectual development. *Developmental Psychology, 19,* 531-543.

Schaie, K.W., & Hertzog, C. (1985). Measurement in the psychology of adulthood and aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 61-92). New York: Van Nostrand Reinhold.

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 and Aging, 13,* 8-20.

Schaie, K.W., & Willis, S.L. (1993). Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging, 8,* 44-55.

Schaie, K.W., Willis, S.L., Hertzog, C., & Schulenberg, J.E. (1987). Effects of cognitive training upon primary mental ability structure. *Psychology and Aging, 2,* 233-242.

Schaie, K.W., Willis, S.L., Jay, G., & Chipuer, H. (1989). Structural invariance of cognitive abilities across the adult life span: A cross-sectional study. *Developmental Psychology, 25,* 652-662.

Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology, 28,* 138-151.

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.

Strother, C.R. , Schaie, K.W., & Horst, P. (1957). The relationship between advanced age and mental abilities. *Journal of Abnormal and Social Psychology, 55,* 166-170.

Thurstone, L.L., & Thurstone, T.G. (1949). *Examiner manual for the SRA Primary Mental Abilities Test (Form 10-14).* Chicago: Science Research Associates.

Tyler, L.E. (1965). *The psychology of human differences* (3rd ed.). New York: Appleton.

Voyer, D., Voyer, S., & Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250-270.

Wiley, D.E., & Wiley, J.A. (1970). The estimation of measurement error in panel data. *American Sociological Review, 35,* 112-117.

Willis, S.L., & Schaie, K.W. (1983). *The Alphanumeric Rotation Test.* Unpublished manuscript, The Pennsylvania State University.

Zelinski, E.M., & Burnight, K.P. (1997). Sixteen-year longitudinal and time-lag changes in memory and cognition in older adults. *Psychology and 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 and Aging, 8,* 176-186.