Psychology and Aging
1998, Vol. 13, No. 1, 8–20

0882-7974/98/$3.00

Longitudinal Invariance of Adult Psychometric Ability Factor Structures Across 7 Years

K. Warner Schaie, Scott B. Maitland, Sherry L. Willis, and Robert C. Intrieri
Pennsylvania State University

The hypothesis that psychometric ability tests retain equivalent factor structures across a 7-year interval was examined in a sample of 984 persons (disaggregated into 6 cohort groups: *M* ages at first test = 32, 46, 53, 60, 67, and 76), assessed in 1984 and 1991 as part of the Seattle Longitudinal Study. A best fitting measurement model was estimated for 20 psychometric tests marking the 6 primary abilities of Inductive Reasoning, Spatial Orientation, Perceptual Speed, Numeric Facility, Verbal Ability, and Verbal Recall. Gender was partialed out at the variable level by including a gender factor. Weak factorial invariance over time was demonstrated for all cohorts. Configural invariance could be demonstrated across all cohort groups. However, weak factorial invariance across groups could be accepted for all but the youngest and oldest groups. Latent means were modeled for the accepted solutions across time and cohort groups.

Much of the literature on the aging of intellectual competence in adulthood has been concerned with the comparison of performance levels between different age groups and within samples as they age across time (cf. Schaie, 1994, 1996a). However, there have also been extensive discussions of the internal validity of such comparisons. These discussions have explicated the theoretical assumptions that should be met in order to demonstrate that the relationship between the measures used to mark the latent theoretical constructs of interest remain invariant across age or time (cf. Horn, 1991; Meredith 1964a, 1964b, 1993).

For the researcher in aging, these issues are of central concern both for longitudinal studies of age change and for age-comparative cross-sectional studies. These concerns are both methodological and substantive. First, the question arises under what circumstances are the measurement assumptions satisfied that permit the comparison of observed variables and the linear composites that estimate their underlying latent constructs across age and time. Second, there is the substantive question whether these measurement assumptions can be met in a particular domain and what the consequences of partial violations of these assumptions might be. Additional substantive questions may also be suggested that involve the test of theoretical propositions related to systematic developmental changes in latent means and factor correlations. (e.g., Nesselroade & Labouvie, 1985; Schaie, 1977, 1988, 1996b; Schaie & Hertzog, 1985).

This study was designed to provide evidence of measurement invariance for groups differing in age in adulthood, both longitudinally and cross-sectionally, for linear composite measures of six important primary ability factors (Ekstrom, French, Harman, & Derman, 1976): Inductive Reasoning (I), Spatial Orientation (S), Perceptual Speed (R), Numerical Facility (N), Verbal Comprehension (V) and Verbal Recall (Ms). It was also designed to test substantive hypotheses about mean differences between and within groups appropriate for the attained level of measurement invariance. Finally, the hypothesis is tested that the factor space of cognitive abilities may change over the life span because of a successive process of cognitive differentiation in younger adulthood and middle age with subsequent dedifferentiation in older adulthood (cf. Reinert, 1970; Werner, 1948).

Measurement Issues

Horn (1991) has argued that when one compares different age groups in terms of means, variances, and correlations for linear composite measures (factor scores), it is necessary to show the invariance of the factor pattern across groups to support the assumption that the common factors measure the same attributes across groups or across time. Horn, McArdle, and Mason (1983) drew attention to an important distinction between two levels of invariance in factor loadings (a distinction first introduced by Thurstone [1947, pp. 360–369]) that may have different implications for age change and age difference research: *configural invariance* and *metric invariance*. Meredith (1993) has spelled out in greater detail what he considers to be necessary conditions to satisfy this factorial invariance at different levels of stringency.

At a minimum it would be expected that the factor pattern across groups or time would display configural invariance. In this case, all measures marking the factors (latent constructs) have their primary non-zero loading on the same ability construct across test occasions or groups. They must also have

K. Warner Schaie, Scott B. Maitland, Sherry L. Willis, and Robert C. Intrieri, Department of Human Development and Family Studies, Pennsylvania State University.

The research reported in this article was supported by National Institute on Aging Grant R37 AG4770.

We gratefully acknowledge the enthusiastic cooperation of members and staff of the Group Health Cooperative of Puget Sound, Seattle, Washington.

Correspondence concerning this article should be addressed to K. Warner Schaie, Department of Human Development and Family Studies, 105 Henderson Building South, Pennsylvania State University, University Park, Pennsylvania 16802. Electronic mail may be sent to kws@psu.edu.

zero loadings on the same measures for all factors. Configural invariance is a minimum condition for factorial invariance. If even this level of invariance is not maintained, then it is likely that developmental processes or cohort effects have resulted in qualitative changes in ability structure. In this case, for example, different numbers or types of constructs may be required at different life stages. No interpretable comparisons would be possible across age or time for variables for which configural invariance cannot be demonstrated.

A second level of factorial invariance (which Meredith, 1993, has termed *weak factorial invariance*) requires the demonstration of equality of the unstandardized factor pattern weights (factor loadings) across groups or time. Several levels of weak factorial invariance have been described. In its weakest form the observed markers not only must have their primary loading on the same ability construct, but the magnitude of the loadings must also be constrained to be proportional across time or between groups. In a stricter case, labeled by Horn (1991) as pattern identity invariance, factor loadings are constrained to be identical across groups or time. The technical and substantive considerations for this level of factorial invariance have found extensive discussion in the literature (cf. Cunningham, 1982; Horn, 1991; Horn & McArdle, 1992; Jöreskog, 1979; Jöreskog & Sörbom, 1989; Meredith, 1964a, 1964b, 1993; Schaie & Hertzog, 1985; Sörbom, 1974, 1975; Thurstone, 1947). If this level of invariance can be accepted, then it becomes possible to test hypotheses about the equivalence of factor means.

Meredith (1993) has further described conditions of what he has termed *strong factorial invariance*. Such conditions involve the equivalence of the unique variances across groups and of the equivalence of intercepts for the mean comparisons.

It seems reasonable to suppose, even given the demonstration of configural invariance, that developmental processes or differential cohort experiences can lead to changes or differences in the magnitude of the regression of the latent constructs on the observed variables. Even though a particular test may measure the same latent construct over different life stages, it may do so with different degrees of efficiency. If the differentiation–dedifferentiation hypothesis is valid, moreover, researchers would then expect an expansion and eventual contraction of factor space, expressed by changing variances and covariances among the latent constructs.

Given the previously provided considerations, it is questionable whether even the assumptions of weak factorial invariance can be met in a complex empirical data set such as is found in many aging studies. In fact, Horn, McArdle, and Mason (1983) early on argued that configural invariance is likely to be the best solution that can be obtained. Nevertheless, it should be possible to demonstrate more stringent levels of invariance for subsystems across some ages and cohorts. Byrne, Shavelson, and Muthén (1989) have proposed therefore that one should test for partial measurement invariance. This proposition has been received with much controversy in the factor-analytic literature. Because of the undue sensitivity of LISREL estimates to local disturbance of model fit, it seems that their position is quite reasonable as seen from the point of view of the substantively oriented scientist.

In any event, it is evident that for both cross-sectional and longitudinal studies, configural invariance remains a minimal requirement, whereas demonstration of some form of metric invariance is essential before valid comparisons of factor scores can be made.

Substantive Background

A number of previous studies in the cognitive aging literature have provided some evidence for the existence of configural invariance in this domain, as well as providing limited data on the tenability of the assumption of metric invariance.

Factorial invariance in longitudinal studies of cognitive abilities over longer time intervals has been examined for a second-order *g* factor for the first five primary mental abilities (Hertzog & Schaie, 1986, 1988). In those studies, stable individual differences were found in the regression of the second-order factor on the primaries over 14-year intervals in three samples with mean ages of 37, 49, and 65 years at the inception of the study.

A short-term repeated measurement study, conducted in the context of cognitive training research (Willis & Schaie, 1986) is also relevant. In that study, five latent ability constructs, each multiply marked, were compared between pre- and posttest. Invariance of factor structure was shown in a sample of 229 participants ranging in age from 62 to 94 years over an interval of several weeks in both control and cognitive intervention groups. This study also shows that factor structure was invariant across gender although there were gender differences in performance levels (Schaie, Willis, Hertzog, & Schulenberg, 1987).

Far more data on factorial invariance across adulthood come from cross-sectional studies. The earliest major analyses were conducted by Cohen (1957) for the standardization sample of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), Cohen found that salient loadings corresponded across groups ranging from younger to older adulthood. However, he also observed substantial shifts in the magnitude of factor loadings, hence, in modern terms, demonstrating configural invariance but not metric invariance.

More recently, Horn and McArdle (1992) were unable to confirm metric invariance across four adult cross-sectional age groups for the Verbal and Performance subscale scores of the WAIS-Revised (WAIS-R; Wechsler, 1981). They also found a significant difference between a two-factor metric model and a corresponding configural model. Both models, however, fit significantly better than a configural one-factor (*g*) model. Hence, Horn and McArdle concluded that there are two factors in the WAIS-R for which both means and covariances are significantly different across age groups across adulthood.

Factorial invariance across age was examined by Cunningham (1981, 1982) for several measures of perceptual speed and verbal ability from the Educational Testing Service (ETS) Kit of Factor Referenced Measures (Ekstrom et al., 1976). For these measures, invariance of factor space was maintained, with quite similar factor loadings, although factor variances and covariances tended to increase with age. Stricker and Rock (1987) in a study of the Graduate Record Examination (GRE) compared factor structures for the first three adult decades and obtained factorial invariance. On the other hand, White and Cunningham (1987) in their comparison of a younger-adult and young-old adult sample had to reject all models that constrained parameters across the two groups and concluded that an additional latent construct was required to explain individual differences in the older group.

We have previously examined the issue of cross-sectional factor invariance in the Seattle Longitudinal Study (SLS; Schaie, Willis, Jay, & Chipuer, 1989). In that analysis we fitted a five-factor model marked by 17 psychometric tests for the 1,621 participants, ages 22–95 years, of the fifth cycle of the SLS. The abilities examined at the latent construct level were Inductive Reasoning, Spatial Orientation, Verbal Ability, Numerical Ability, and Perceptual Speed. The overall model had a good fit, $\chi^2(107, N = 1621) = 946.62$; GFI = .936. Participants were then disaggregated into eight subsets with mean ages 29, 39, 46, 53, 60, 67, 81 and 90 years. Each set was fitted separately to the accepted overall model, under conditions of configural invariance, metric invariance, and constraining the factor variance–covariance matrices. Metric invariance was obtained for all subsets except the 90-year-old group. However, relaxing the equivalence of the variance–covariance matrices significantly improved fit for all cohorts, and relaxing the equivalence of factor-loading requirement resulted in further significant improvement in fit, except for the 39-year-old and 67-year-old groups. In a reanalysis of this data set (Schaie, Dutta, & Willis, 1991) a Verbal Recall factor was added, such that 6 latent constructs were measured with 20 observed markers. The fit for this model was $\chi^2(151, N = 1628) = 1{,}444.26$, $p < .001$; GFI = .934.

The Study

During the sixth-wave SLS data collection in 1991 we were able to retest 982 individuals who had taken the entire SLS battery in 1984. Hence, we are finally able to report data on factorial invariance of psychometric ability structures in samples that have been followed longitudinally over a 7-year period and to examine longitudinal and cross-sectional factor invariance, including an analysis of mean structures, in the same population. We test the hypothesis that covariance structures show more stringent conditions of invariance in longitudinal (within cohorts) comparisons than in cross-sectional (between cohort) comparisons. Specifically, the hypothesis is tested that metric invariance (weak factorial invariance) can be shown to be the most parsimonious model in longitudinal, but not necessarily in cross-sectional, comparisons. We also test the proposition that latent factor scores differ across cohort groups, but remain stable within groups, except in advanced old age. Finally, we test the hypotheses that there is expansion and contraction of the factor space across the adult portion of the life span and that factor pattern invariance cannot be demonstrated at advanced ages.

Method

Participants

Our inquiry into adult cognitive functioning began some 42 years ago when we randomly sampled 500 participants equally distributed by sex and age across the range from 20 to 70 years from the approximately 18,000 individuals who were then members of a health maintenance organization in the Pacific Northwest (Schaie, 1958, 1994, 1996b). The survivors of the original sample were retested and additional panels were added in 7-year intervals. The sampling frame represents a broad distribution of educational and occupational levels, covering the upper 75% of the socioeconomic spectrum. The population frame from which we have been sampling repeatedly has grown to a membership of over 400,000 individuals, but the general demographic characteristics of the sampling frame remain quite comparable.

The data to be examined here include the 982 community-dwelling individuals (442 men and 540 women) who were examined in both the fifth SLS cycle during 1983–1985 and the sixth cycle in 1990–1992, with an average time interval of 7 years. For purposes of cohort comparisons, we divided the sample into six age cohorts with mean ages 32, 46, 53, 60, 67, and 76 years at first test, respectively (see Table 1 for a breakdown by age–cohort). These groups represent 7-year age ranges, except for the groups aged 32 and 76 years, where adjacent age ranges were collapsed to produce sample sizes sufficient for multivariate analyses. The entire sample tested at Time 1 included 1,608 individuals. Hence, there was a 39% attrition rate. Attrition was greatest for the groups aged 32 and 76 years because of death and disability in the older participants and because of job mobility in the youngest group.

The individuals included in this analyses had an average educational level of 14.7 years (SD = 3.0; range = 1–20 years); their family income averaged $27,604 ($SD$ = $8,178; range = $1,000–$50,000+). Occupational levels were rated on a scale from 0 for unskilled to 9 for professional occupations. Those individuals gainfully employed at the time of assessment averaged an occupational level of 6.9 (SD = 1.87). Most frequent occupations represented involve skilled trades, clerical sales, managerial, and semiprofessional jobs (see Schaie, 1996b, for more details).

Measurement Variables

The original SLS psychometric ability battery was expanded to permit structural analyses that require multiple measures to mark each ability factor. All tests are slightly speeded to be suitable for group administration. The longitudinal markers included in this battery of necessity (i.e., for consistency across successive test administrations) employ the test booklet and answer sheet format used since the beginning of the SLS (Thurstone & Thurstone, 1949). However, print size on answer sheets has been enlarged from the original. All other forms use disposable booklets with suitably enlarged type upon which answers are marked directly (cf. Ekstrom et al., 1976; Schaie, 1985). Table 2 lists 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 (Schaie, Willis, Hertzog, & Schulenberg, 1987). A brief description of the primary mental abilities (PMA) and the measures marking them is given next:

Table 1
Subsamples Entering the Multigroup Analyses

Group	SLS cohort	Dates of birth	*n* M	*n* F	*n* *N*	Mean age at Time 1
1	2–4	1897–1913	69	93	162	76
2	5	1914–1920	85	109	194	67
3	6	1921–1927	93	90	183	60
4	7	1928–1934	61	86	147	53
5	8	1935–1941	54	72	126	46
6	9–11	1942–1962	80	90	170	32
Total sample			442	540	982	59

Note. Following the convention used in all reports from the Seattle Longitudinal Study (SLS), lower cohort numbers represent earlier-born (older) participants. M = male; F = female.

Table 2
Psychometric Intelligence Measurement Battery

Primary ability	Test	Source	Test–retest correlation
Inductive Reasoning	PMA Reasoning (1948)	Thurstone & Thurstone, 1949	.884
	ADEPT Letter Series (Form A)	Blieszner et al., 1981	.839
	Word Series	Schaie, 1985	.852
	ETS Number Series	Ekstrom et al., 1976	.833
Spatial Orientation	PMA Space (1948)	Thurstone & Thurstone, 1949	.817
	Object Rotation	Schaie, 1985	.861
	Alphanumeric Rotation	Willis & Schaie, 1983	.820
	ETS Cube Comparisons	Ekstrom et al., 1976	.951
Numerical Facility	PMA Number (1948)	Thurstone & Thurstone, 1949	.875
	ETS Addition (N-1)	Ekstrom et al., 1976	.937
	ETS Subtraction & Multiplication (N-3)	Ekstrom et al., 1976	.943
Verbal Comprehension	PMA Verbal Meaning (1948)	Thurstone & Thurstone, 1949	.890
	ETS Vocabulary (V-2)	Ekstrom et al., 1976	.928
	ETS Advanced Vocabulary (V-4)	Ekstrom et al., 1976	.954
Perceptual Speed	ETS Identical Pictures	Ekstrom et al., 1976	.814
	ETS Finding As	Ekstrom et al., 1976	.860
	ETS Number Comparison	Ekstrom et al., 1976	.865
Verbal Recall	Immediate Recall	Zelinski et al., 1993	.820
	Delayed Recall	Zelinski et al., 1993	.732
	PMA Word Fluency	Thurstone & Thurstone, 1949	.896

Note. PMA = Primary Mental Abilities test; ADEPT = Adult Development and Enrichment Project; ETS = Educational Testing Service.

Inductive Reasoning (I)

This is the ability to identify novel aspects of relationships, and to infer principles or rules from observing the regular occurrence of instances or relationships.

PMA Reasoning. The participant is shown a series of letters (e.g., a b c b a d e f e) and is asked to identify the next letter in the series.

ADEPT Letter Series. The Adult Development and Enrichment Project (ADEPT) Letter Series is a parallel form to the PMA Reasoning test.

Word Series. The participant is shown a series of words (e.g., January, March, May) and is asked to identify the next word in the series. Positional patterns used in this test are identical to the PMA Reasoning test.

Number Series. The participant is shown a series of numbers (e.g., 6, 11, 15, 18, 20) and is asked to identify the next number that would continue the series.

Spatial Orientation (S)

The ability to visualize and mentally manipulate spatial configurations, to maintain orientation with respect to spatial objects, and to perceive relationships among objects in space.

PMA Space. The study participant is shown an abstract figure and is asked to identify which of six other drawings represents the model in two-dimensional space.

Object Rotation. The participant is shown a line drawing of a meaningful object (e.g., an umbrella) and is asked to identify which of six other drawings represents the model rotated in two-dimensional space.

Alphanumeric Rotation. The participant is shown a letter or number and is asked to identify which of six other drawings represents the model rotated in two-dimensional space.

Test stimuli in the Object and Alphanumeric Rotation tests have the same angle of rotation as the abstract figures in the PMA Space test.

Cube Comparisons. In each item, two drawings of a cube are presented; the participant is asked to indicate whether the two drawings are of the same cube, rotated in three-dimensional space.

Numerical Facility (N)

The ability to understand numerical relationships and compute simple arithmetic functions.

PMA Number. The participant checks whether additions of simple sums shown are correct or incorrect.

Addition. This is a test of speed and accuracy in adding three single- or two-digit numbers.

Subtraction and Multiplication. This is a test of speed and accuracy with alternate rows of simple subtraction and multiplication problems.

Verbal Comprehension (V)

Language knowledge and comprehension is measured by assessing the scope of a person's recognition vocabulary.

PMA Verbal Meaning. A four-choice synonym test that is highly speeded.

ETS Vocabulary II. A five-choice synonym test.

ETS Vocabulary IV. A five-choice synonym test consisting mainly of difficult items.

Perceptual Speed (P)

The ability to find figures, make comparisons, and carry out other simple tasks involving visual perception, with speed and accuracy.

Identical Pictures. The participant identifies which of five numbered shapes or pictures in a row are identical to the model at the left of the row.

Finding As. In each column of 40 words, the participant must identify the five words containing the letter *a*.

Number Comparison. The participant inspects pairs of multidigit numbers and indicates whether the two numbers in each pair are the same or different.

Verbal Recall (Ms)

The ability to encode, store and recall meaningful language units.

Immediate Recall. Participants study a list of 20 words for 3.5 min.

They are then given an equal period of time to recall the words in any order.

Delayed Recall. Participants are asked to recall the same list of words as in Immediate Recall after an hour of intervening activities (other psychometric tests).

PMA Word Fluency. The participant freely recalls as many words as possible according to a lexical rule within a 5-min period.

Assessment Procedure

The measures described previously were administered to small groups of participants as part of a broader 5-hr battery spread over two sessions. The tests were administered by an examiner who was assisted by a proctor. Testing locations were at familiar sites close to the homes of our participants.

Statistical Procedures

Longitudinal confirmatory factor analyses were conducted to assess two levels of invariance: configural invariance and weak factorial invariance. The initial structural model for this analysis (i.e., specification of the salient and zero-loadings factor pattern) was based on a slight modification of the model reported by Schaie, Dutta, and Willis (1991). In particular we added a gender variable and a gender factor to account for variance in the structural model attributable to gender differences.

We extended this factor structure to a repeated measures multigroup factor model for panel data (Alwin, 1988; Jöreskog, 1979). Analyses were conducted on the variance–covariance matrix, with results rescaled into correlation metric for ease of interpretation. The factor pattern was identified by fixing the highest loading on each factor to 1.0 in the pattern matrix. A 41 × 41 covariance matrix of Time 1 (1984) and Time 2 (1991) scores contained the same 20 cognitive measures plus a gender variable, comprising six factors for each occasion plus a gender factor. The model accounted for autocorrelation of the residuals because repeated measures of the same scales were used (Sörbom, 1975; Wiley & Wiley, 1970). The covariances of the residuals are orthogonal to the common factor covariances over time and are needed to provide unbiased estimates of the stability of individual differences in the factors (see Hertzog & Schaie, 1986; Sörbom, 1975).

Five versions of the repeated measures factor model were tested: The first model tests the invariance of factor patterns for the entire model (across time and cohorts). This is the reference configural invariance model (M1). Next, we test four alternative weak factorial invariance models, all of which are nested in the configural invariance model. The first of these constrains the factor-loadings invariant across time (longitudinal invariance, M2). The second model frees the longitudinal loadings, but constrains loadings across cohorts (replicated cross-sectional invariance, M3). The third model constrains loadings over both cohort and time (longitudinal and cross-sectional invariance, M4). The fifth model is a partial form of M4. It constrains factor loading across time for all groups, but constrains loading across cohorts for all but the youngest and oldest cohorts (longitudinal and partial cross-sectional invariance, M5).

Analyses of Mean Structures

These analyses followed the work of Jöreskog and Sörbom (1989), Mandys, Dolan, and Molenaar (1994) and Sörbom (1974), using covariance and mean matrices. The procedure involved first fitting the covariance matrices to an acceptable model, then extending the models to include the mean structures.

Mean deviations were investigated by two methods. First, the means for cohort groups were examined for change between 1984 and 1991. This approach is analogous to a within-subjects analysis in multivariate analysis of Variance (MANOVA). The weak factorial invariance model was used with factor-loading estimates held invariant over time. Means structures were tested by allowing the intercepts to be freely estimated in 1984 and fixing the 1991 intercepts equal to these values. Latent mean parameters for 1984 were fixed to zero but allowed to be freely estimated for the same six latent factors in 1991. This strategy yielded deviation values from Time 1 (1984) to Time 2 (1991) and are the basis for estimating change or stability of latent means structures within groups over time. Second, between-groups deviation values on the 12 estimated means was tested in the following way. Six mean values were fixed to zero in a reference group; Cohort Group 5, the youngest group for which cross-cohort invariance could be accepted, was used for this purpose. Intercepts from the reference group were then tested for fit and mean deviation values were estimated.

Tests of Model Fit

Model fit was assessed using LISREL 8 (Jöreskog & Sörbom, 1989). LISREL 8 generates a covariance matrix using the specific factor structures as a reference. The factor structure is acceptable if only small discrepancies exist between the actual and estimated matrices. Because χ^2 alone is not an adequate measure of model fit when a relatively large number of participants are used (Marsh, Balla, & McDonald, 1988), several other criteria were also used to evaluate model fit. These included the normed fit index and the nonnormed fit index (NFI, NNFI; Bentler & Bonnet, 1980); the comparative fit index (CFI; Bentler, 1990); the root-mean-square error of approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980); the goodness-of-fit index (GFI; Jöreskog & Sörbom, 1989); the *Z* ratio (χ^2/df; Bollen, 1989); expected cross-validation index (ECVI) and estimated noncentrality parameters (NCP; Browne & Cudeck, 1989, 1993).

Comparisons of fit for the nested models employ the $\Delta\chi^2$ index (Jöreskog & Sörbom, 1989), which is distributed as χ^2. We accept a chi-square difference significant at or beyond the 1% level of confidence as indicating a significant decrease in model fit for the more constrained model.

Results

We first report findings on the measurement models. The first of these models tests the fit for the configural invariance (i.e., equality of the factor pattern of salient and zero loadings). We next report the weak factorial invariance models. A strong factorial invariance model is then examined to include intercepts and factor means. In the text, we report model fit indices in terms of χ^2, GFIs, and *Z* ratios. Table 3, in addition to these basic indices, lists values for the alternative indices RMSEA, NFI, NNFI, CFI, ECVI, and NCP; see previous section). Next we report analyses of latent mean differences for the accepted factor model across time and groups. Finally, substantive results are given with respect to factor-loading differences between cohorts and cohort-specific change over time in the factor correlations.

Measurement Invariance

Configural Invariance

We first established a baseline model (M1) that demonstrates factor pattern invariance across time group, the minimal condition necessary for any comparisons whether they involve cross-sectional or longitudinal data. In this, as in subsequent analyses, the factor variance–covariance matrices (ψ) and the unique variances ($\theta\epsilon$) are allowed to be freely estimated across time

Table 3
Comparisons of Measurement Models

Multigroup models	*dfs*	χ^2 (N = 982)	*p*	GFI	RMSEA	NFI	NNFI	CFI	ECVI	NCP	*Z* ratio
Configural											
Factor pattern (M1)	3,888	5,155.71	.001	.81	.018	.87	.95	.96	7.90	1,267.71	1.33
Weak invariance											
Time Invariant LY (M2)	3,990	5,288.98	.001	.81	.018	.87	.95	.96	7.83	1,298.98	1.33
Cohort Invariant LY (M3)	4,258	5,790.65	.001	.78	.019	.85	.95	.96	7.79	1,532.65	1.36
Time + Cohort Invariant LY (M4)	4,275	5,801.09	.001	.78	.019	.85	.95	.96	7.77	1,526.09	1.36
Cohorts 2–5 Invariant, Cohorts 1 and 6, time invariant only (M5)[a]	4,161	5,484.20	.001	.81	.018	.86	.95	.96	7.68	1,323.20	1.32
Strong invariance											
M5/Intercepts (M6)	4,161	5,484.20	.001	.81	.018	.86	.95	.96	8.18	1,323.20	1.32
M5/Change over Time (M7)	4,245	5,713.87	.001	.81	.019	.86	.95	.96	8.25	1,468.87	1.35
M5/Cohort Differences (M8)	4,301	6,015.13	.001	.80	.020	.85	.94	.95	8.44	1,714.13	1.40

Note. GFI = goodness of fit index; RMSEA = root mean square error of approximation; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; ECVI = expected cross-validation index; NCP = estimated noncentrality parameter; *Z* ratio = χ^2/df; LY = lambda Y.
[a] Accepted model.

and groups. The model differs from that reported by Schaie et al. (1991) by setting the Word Fluency parameter on Verbal Recall to zero and by adding a gender factor that allows salient loadings for all variables. The variances ($\theta\epsilon$) were allowed to be free across time but constrained across group. Given the complexity of this data set, this model shows a reasonably good fit: $\chi^2(3888, N = 982) = 5{,}155.71, p < .001$ (GFI = .81; *Z* ratio = 1.33).

Weak Factorial Invariance

Four weak factorial invariance models were tested, the first two of which are nested in M1. The first model (M2) constrains the factor loadings (λ) equal across time. This is the critical test for the invariance of factor loading within a longitudinal data set. This model resulted in a slight but statistically nonsignificant reduction in fit: $\chi^2(3990, N = 982) = 5{,}288.98, p < .001$ (GFI = .81; *Z* ratio = 1.33); $\Delta\chi^2(102) = 133.27$ $p > .01$. Hence, we conclude that we can accept invariance within groups across time.

A second model (M3) allowed the values of the factor loadings (λ) to be free across time, but constrained to be equal across cohort groups. This is the test of factorial invariance for the replicated cross-sectional comparisons across cohorts. The model showed a highly significant reduction in fit, as compared to M1: $\chi^2(4258, N = 982) = 5{,}790.65, p < .001$ (GFI = .78; *Z* ratio = 1.36); $\Delta\chi^2(370) = 634.94, p < .001$. As a consequence this model must be rejected, and we conclude that there are significant differences in factor loadings across cohorts.

The third model (M4), which is nested in both M2 and M3, constrains the factor loadings (λ) to be equal across time and group. This particular model, if accepted, would demonstrate factorial invariance both within and across groups. The fit for this model was: $\chi^2(4275, N = 982) = 5{,}801.09$ $p < .001$ (GFI = .78; *Z* ratio = 1.36). The reduction in fit is significant in the comparison with M2, $\Delta\chi^2(285) = 512.11, p < .001$; but not significant when compared with M3, $\Delta\chi^2(17) = 10.44, p > .01$. This model again confirms that we can accept time invariance within cohorts but cannot accept invariance across cohort groups.

Before totally rejecting factorial invariance across all cohorts, we also tested a partial invariance model that constrains factor loadings across time and constrains factor loadings across all but the youngest and oldest cohorts (M5). We arrived at this model by examining confidence intervals around individual factor loadings in M2. This model is nested within M2. The fit for this model was: $\chi^2(4161, N = 982) = 5{,}484.20, p < .001$ (GFI = .81; *Z* ratio = 1.32). The reduction in fit in the comparison with M2 is not significant $\Delta\chi^2(171) = 195.22, p < .09$. Hence, we conclude that this model can be accepted and that we have demonstrated partial invariance across cohorts.

Because we accept Model M5 (invariance across time and partial invariance across groups), we report time-invariant factor loadings separately for Cohorts 1 and 6, as well as a set of loadings for Cohorts 2–5 in Table 4. Note that the factor intercorrelations are reported in Table 5 for both times of measurement because they are not constrained across time. (See next section for substantive analyses of differences between factor intercorrelations.)

When differences in factor loadings between cohorts are examined, it is found that the significant cohort differences are quite localized. No significant differences were found on the Verbal Recall factor. Significant differences on Inductive Reasoning were found for all markers except PMA Reasoning. Significant differences on Spatial Orientation were found for all but the Object Rotation test. Loadings increase with age for Alphanumeric Rotation, but decrease for Cube Comparison. On Perceptual Speed, loadings for Number Comparison decrease and loadings on Verbal Meaning increase with age. The loading of Verbal Meaning on the Verbal Comprehension factor, by contrast, decreases with age. Finally, there is a significant increase for the loadings of Number Comparison on the Numerical Facility Factor.

Strong Factorial Invariance

Given the acceptance of M5, we next estimated a strong factorial invariance model that constrains factor loadings over time and groups but freely estimates the intercepts for the means

Table 4
Rescaled Solution for Multigroup Analyses: Factor Loadings

Factors/variables	Cohort 1	Cohorts 2–5	Cohort 6
Inductive Reasoning			
PMA Reasoning	.88	.88	.88
ADEPT Letter Series	.67[ab]	.85	.98[a]
Word Series	.94[a]	.78	.88
Number Series	.62[ab]	.72	.73
Spatial Orientation			
PMA Space	.60[ab]	.78	.99[a]
Object Rotation	.82	.82	.82
Alphanumeric Rotation	.79[b]	.79	.52[a]
Cube Comparison	.24[ab]	.44	.78[a]
Verbal Comprehension			
PMA Verbal Meaning	.36[ab]	.54	.55
ETS Vocabulary	.94[ab]	.86	.84
Advanced Vocabulary	.92	.92	.92
Word Fluency	.30	.39	.38
Numerical Facility			
PMA Number	.84	.86	.79[a]
Addition	.90	.98	.88[a]
Subtraction and Multiplication	.86	.86	.86
Number Comparison	.37[ab]	.20	.12[a]
Perceptual Speed			
Identical Pictures	.61	.61	.61
Number Comparison	.34[b]	.49	.78[a]
Finding A's	.53	.51	.72
Word Fluency	.24	.32	.37
PMA Verbal Meaning	.68[ab]	.44	.28
Recall Memory			
Immediate Recall	.99	.92	.90
Delayed Recall	.87	.87	.87

[a] Differs significantly ($p < .01$) from Cohorts 2–5. [b] Differs significantly ($p < .01$) from Cohort 6 (youngest cohort).

structure (M6). Model M6, of course, yields the same fits as M5. We then estimate the model by also constraining the intercepts across time within cohorts (providing latent mean deviations; M7). These deviations provide estimates of latent mean change within cohorts over the 7-year period. The fit of the model extended to include the within cohort means structure (M7) was $\chi^2(4245, N = 982) = 5{,}713.87$ $p < .001$ (GFI = .81; Z ratio = 1.35).

The final measurement model freely estimates the intercepts for the reference group (Cohort 5) while setting all means for this group to zero. The intercepts from the reference group are then fitted to the remaining groups and group difference values are estimated. The fit of the model extended to include the cohort comparisons (M8) was $\chi^2(4301, N = 982) = 6{,}015.13$ $p < .001$ (GFI = .80; Z ratio = 1.40).

Gender-partitioned values (our method of controlling for gender differences) are reported in Table 6. These values indicate gender differences in the observed variables. There are relatively few significant values in this table. Differences are in favor of men for Number Series, Cube Comparison, and PMA Space. Differences in favor of women occur for Number Comparison and the Verbal Recall measures.

Tests of Substantive Hypotheses

Dedifferentiation of Factor Space

This data set also allows a substantive test of the proposition that factor variances decrease and factor covariance increase with age during adulthood. This is essentially an operationalization of the differentiation–dedifferentiation hypothesis. This hypothesis can be tested by examining models that constrain the factor variance–covariance matrix (ψ) across time within cohorts as well as across successive cohort groups.

Consistent with the tests of measurement models described earlier, we first add the time constraint for the factor variance–covariance matrices to our accepted measurement model (M5). The fit of this model (M9) was $\chi^2(4287, N = 982) = 5{,}642.74$, $p < .001$ (GFI = .80; Z ratio = 1.32). The reduction in fit in the comparison with M5 is not significant $\Delta\chi^2(126) = 158.51$, $p < .033$. Hence, we conclude that this model can be accepted and that we fail to confirm dedifferentiation over a 7-year time period.

Table 5
Rescaled Solution for Multigroup Analyses: Factor Intercorrelations

	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 5		Cohort 6	
Factors	1984	1991	1984	1991	1984	1991	1984	1991	1984	1991	1984	1991
Inductive Reasoning/Spatial Orientation	.52	.70	.58	.65	.62	.63	.55	.53	.59	.52	.56	.53
Inductive Reasoning/Verbal Comprehension	.66	.66	.64	.62	.61	.68	.45	.45	.27	.26	.49	.45
Inductive Reasoning/Numerical Facility	.47	.63	.55	.51	.47	.52	.39	.36	.59	.50	.45	.52
Inductive Reasoning/Perceptual Speed	.75	.86	.76	.77	.76	.67	.64	.60	.85	.70	.63	.70
Inductive Reasoning/Verbal Recall	.45	.51	.53	.61	.36	.36	.32	.37	.23	.27	.27	.24
Spatial Orientation/Verbal Comprehension	.31	.43	.28	.30	.27	.38	.22	.14	.07	.01	.10	.05
Spatial Orientation/Numerical Facility	.23	.46	.29	.39	.36	.38	.23	.24	.31	.33	.15	.23
Spatial Orientation/Perceptual Speed	.69	.78	.54	.71	.61	.57	.51	.52	.51	.56	.36	.48
Spatial Orientation/Verbal Recall	.28	.44	.27	.40	.02	.13	.13	.12	.23	.20	.14	.10
Verbal Comprehension/Numerical Facility	.35	.44	.32	.31	.25	.26	.09	.14	.06	.03	.22	.19
Verbal Comprehension/Perceptual Speed	.44	.52	.28	.37	.50	.40	.26	.31	.21	.15	.33	.24
Verbal Comprehension/Verbal Recall	.44	.51	.44	.53	.54	.50	.28	.36	.41	.43	.32	.30
Numerical Facility/Perceptual Speed	.67	.79	.63	.71	.59	.60	.50	.51	.77	.63	.67	.62
Numerical Facility/Verbal Recall	.23	.38	.34	.43	.15	.15	.09	.15	.10	.16	.13	.03
Perceptual Speed/Verbal Recall	.44	.56	.41	.56	.35	.34	.20	.31	.45	.36	.51	.35

Table 6
Gender Partitioned Values for Multigroup Analyses

	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 5		Cohort 6	
Factors	1984	1991	1984	1991	1984	1991	1984	1991	1984	1991	1984	1991
PMA Reasoning	.10	.16	.15	.16	−.01	−.05	.13	.12	.07	.06	.02	.03
ADEPT Letter Series	.00	.04	.04	.07	−.11	−.08	.03	.01	−.06	.04	−.05	.05
Word Series	.17	.18	.08	.14	.05	−.01	.14	.12	.07	.06	.06	.08
Number Series	−.11	−.09	−.19[a]	−.15	−.30[a]	−.28[a]	−.24[a]	−.22[a]	−.23[a]	−.25[a]	−.27[a]	−.31[a]
PMA Space	−.29[a]	−.12	−.24[a]	−.16	−.33[a]	−.39[a]	−.15	−.17	−.22	−.23	−.11	−.25[a]
Object Rotation	−.23[a]	−.07	−.20[a]	−.25[a]	−.21[a]	−.38[a]	−.13	−.13	−.21	−.21	−.19	−.21[a]
Alphanumeric Rotation	−.12	−.04	−.13	−.15	−.05	−.08	−.14	−.10	−.06	−.09	−.12	−.15
Cube Comparison	−.07	−.04	−.20[a]	−.24[a]	−.37[a]	−.34[a]	−.36[a]	−.38[a]	−.36[a]	−.34[a]	−.30[a]	−.20[a]
PMA Verbal Meaning	.06	.05	.03	.07	−.04	−.05	.06	.10	.03	.12	.01	.02
ETS Vocabulary	−.01	.10	.02	−.01	−.10	−.14	.07	.14	.15	.08	−.05	−.07
Advanced Vocabulary	−.03	.02	−.03	.01	−.17	−.16	−.01	.04	.03	.03	−.14	−.12
PMA Number	−.18	.00	−.05	−.04	−.29[a]	−.27[a]	−.06	−.01	−.12	.01	−.05	−.15
Addition	−.12	.02	.10	.11	−.13	−.14	.08	.08	−.03	.03	.01	−.02
Subtraction/Multiplication	−.12	−.02	.04	.07	−.14	−.14	.06	.02	−.13	−.11	−.03	−.06
Identical Pictures	.05	.16	.01	.01	.00	.01	−.05	−.02	.00	−.06	−.03	.01
Number Comparison	.14	.14	.20[a]	.14	.13	.12	.20	.19	.15	.07	.23[a]	.22[a]
Finding As	.14	.09	.11	.17	.13	.09	.24[a]	.17	.10	.12	.10	.11
Word Fluency	.21[a]	.22[a]	.04	.06	.01	.06	.07	.09	.20	.18	.01	.09
Immediate Recall	.27[a]	.24[a]	.23[a]	.22[a]	.08	.11	.15	.27[a]	.22	.12	.15	.19
Delayed Recall	.26[a]	.25[a]	.28[a]	.24[a]	.14	.17	.18	.29[a]	.17	.15	.18	.25[a]

[a] Indicates significant gender difference for observed variable ($p < .01$). Negative values favor men, positive values favor women.

We next constrain the factor variance–covariance matrices across groups. This is a test of the proposition of dedifferentiation occurring as a function of cohort membership. The fit of this model (M10) was $\chi^2(4556, N = 982) = 6{,}041.64$, $p < .001$ (GFI = .79; Z ratio = 1.33). The reduction in fit in the comparison with M5 is significant $\Delta\chi^2(374) = 453.33$, $p < .01$. This model cannot be accepted, and thus we conclude that there are indeed significant difference in the variance–covariance matrices across cohorts.

For completeness, we also tested a model that constrained ψ across both time and cohorts. This models nests in M9. The fit of this model (M11) was $\chi^2(4577, N = 982) = 6{,}081.27$, $p < .001$ (GFI = .79; Z ratio = 1.33). The reduction in fit in the comparison with M9 is significant $\Delta\chi^2(290) = 438.53$, $p < .01$. As expected, this model cannot be accepted.

Consistent with our earlier strategy in considering partial invariance models, we also test the analogy of M5 for the variance–covariance differences. That is, we constrain the ψ matrices across time and for Cohorts 2–5, but allow the ψs for Cohorts 1 and 6 to differ. This model is nested within M9. The fit of this model (M12) was $\chi^2(4535, N = 982) = 5{,}938.53$, $p < .001$ (GFI = .79; Z ratio = 1.31). The reduction in fit in the comparison with M9 is not significant $\Delta\chi^2(248) = 295.79$, $p < .02$. This model can be accepted, and we conclude that there are significant differences in the factor intercorrelations across some but not all cohorts. Table 7 shows complete fit statistics for this model. Table 8 provides the factor intercorrelations for the accepted model.

Examination of the factor intercorrelations across cohort groups provides some evidence for the dedifferentiation of factor structures. All intercorrelations for the oldest cohort are numerically larger for the oldest cohort, and smaller differences in the same direction occur for the middle-aged as against the younger-adult cohort. However, none of the middle-aged and younger-adult values differ significantly. The significant differences for the old-old cohort (Cohort 1) seem to affect only the intercorrelations involving the Spatial Orientation factor, except for the relation with Verbal Comprehension.

Factor Mean Differences

Longitudinal changes. The final set of substantive hypotheses involves comparison of the latent means estimated for the accepted measurement models. First, intercepts for Time 1 (1984) were estimated and held invariant for Time 2 (1991). Estimated 1991 latent means are represented as deviations from 1984 to 1991. This allows the test of mean deviations across time. Table 9 provides mean deviation values and standard errors. Mean deviation values were statistically significant at the .01 level for approximately two thirds (23/36) of the values obtained.

Figure 1 provides a plot of these difference values for each cohort group. The graph illustrates a similar trend for each group. It is noteworthy that the oldest two cohort groups (from age 67 to 74 and age 76 to 83) declined significantly on all six cognitive ability factors. Cohort 3 (from age 60 to 67) declined on all ability factors except Verbal Recall. Cohort 4 (from age 53 to 60) remained stable on five of the six cognitive abilities with statistically significantly decline evidenced only for Numeric Facility. Cohort 5 (from age 46 to 53) was stable on four of the six abilities. It showed improved performance for the Verbal Comprehension and Verbal Recall factors. The youngest group, Cohort 6 (from age 32 to 39), remained stable on Numeric Facility, Perceptual Speed, and Verbal Recall while showing significant improvement in performance on Inductive Reasoning, Spatial Orientation, and Verbal Comprehension. All

Table 7
Comparisons of Substantive Models

Multigroup models	*dfs*	χ^2 ($N = 982$)	*p*	GFI	RMSEA	NFI	NNFI	CFI	ECVI	NCP	*Z* ratio
M5 + ψ Time Invariant (M9)	4287	5,642.74	.001	.80	.018	.86	.96	.96	7.58	1,355.74	1.32
M5 + ψ Cohort Invariant (M10)	4556	6,041.64	.001	.79	.018	.85	.95	.96	7.44	1,485.64	1.33
M5 + ψ time + cohort invariant (M11)	4577	6,081.27	.001	.79	.018	.85	.95	.96	7.44	1,504.27	1.33
M5 with ψ invariant Cohorts 2–5, time invariant only Cohorts 1, 6 (M12)[a]	4535	5,938.53	.001	.79	.018	.85	.96	.96	7.38	1,403.53	1.31

Note. GFI = goodness of fit index; RMSEA = root mean square error of approximation; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; ECVI = expected cross validation index; NCP = estimated noncentrality parameter; *Z* ratio = χ^2/df; ψ = factor variance–covariance matrices.
[a] Accepted model.

groups obtained lower estimated mean deviation scores in 1991 on the Numeric Facility factor, although the across-time differences for Cohorts 5 and 6 were not statistically significant.

Cross-sectional differences. Factor mean differences between cohorts were tested by setting all estimated cognitive ability factors for 1984 and 1991 at zero for Cohort 5, the youngest of the groups across which invariance could be established. The intercepts for Cohort 5 were estimated and tested for invariance for Cohorts 1–4 and 6. The designation of the reference was arbitrary, and any of the six groups could have been selected. This analysis provides a way to judge differences from a reference group. Table 10 provides mean deviation scores and standard errors for the between-groups differences model.

Table 8
Factor Intercorrelations for Cohort Groups

Factors	Cohort 1	Cohorts 2–5	Cohort 6
Inductive Reasoning/Spatial Orientation	.64	.58	.57
Inductive Reasoning/Verbal Comprehension	.59	.54	.53
Inductive Reasoning/Numerical Facility	.55	.50	.47
Inductive Reasoning/Perceptual Speed	.79	.73	.71
Inductive Reasoning/Verbal Recall	.44	.39	.38
Spatial Orientation/Verbal Comprehension	.27	.22	.23
Spatial Orientation/Numerical Facility	.38[a]	.30	.28
Spatial Orientation/Perceptual Speed	.71[a]	.59	.55
Spatial Orientation/Verbal Recall	.29[a]	.21	.18
Verbal Comprehension/Numerical Facility	.27	.23	.23
Verbal Comprehension/Perceptual Speed	.40	.34	.34
Verbal Comprehension/Verbal Recall	.48	.42	.41
Numerical Facility/Perceptual Speed	.72	.63	.62
Numerical Facility/Verbal Recall	.26	.20	.19
Perceptual Speed/Verbal Recall	.45	.40	.42

[a] Differs significantly ($p < .01$) from Cohort 6 (youngest cohort).

The *t* values were statistically significant for 62% (37/60) of the 60 possible tests. Those not statistically significant were located on Verbal Comprehension (all in 1984 and 1991) and Numeric Facility (all in 1984 and five in 1991), as well as one value for all other factors (Cohort 4) in 1984. Figure 2 provides a plot of each group's 1984 and 1991 estimated mean deviation scores displayed as deviation values from the reference group. The resulting profiles are remarkably similar. It is particularly noteworthy that estimated latent means for Verbal Comprehension and Numerical Facility in 1984 and 1991 display marked stability across groups and show limited variability from the reference group.

Discussion

This article provides empirical data with respect to one of the most fundamental assumptions in the study of developmental processes. This assumption refers to the equivalence of the constructs that are to be compared across groups of different ages or within groups of individuals at different ages measured over time. Specifically, this includes the assumption that the regressions of the observed variables on the constructs of interest (factor loadings) remain invariant across these comparisons.

We defined measurement invariance at three levels of stringency: configural invariance (identical patterns of salient and nonsalient loadings), four levels of weak factorial invariance (equivalence of factor loadings), and four levels of strong factorial invariance (equivalence of factor loadings and intercepts of mean structures). Demonstration of configural invariance is essential if we want to be certain that we are comparing the same constructs across time and groups. If we wish to show that there is a constant relationship between these constructs and the observations by which they are measured, we must, as a minimum, provide evidence of weak factorial invariance. Demonstration of strong factorial invariance, moreover, allows us to be certain that comparisons of latent factor means are meaningful.

In this study we tested the previously discussed levels of factorial invariance within cohorts (longitudinally) over a 7-year period for a sample of adults ranging in age at initial test from 22 to 87 years, divided into six subsets with more limited age ranges. We did so within the domain of psychometric intelligence, as sampled by multiple markers of the six latent con-

Table 9
Mean Deviation Values and Standard Errors for Within-Group Change

	Factors											
	Inductive Reasoning		Spatial Orientation		Verbal Comprehension		Numeric Facility		Perceptual Speed		Verbal Recall	
Group	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Cohort 1	−1.89[a]	.23	−4.94[a]	.66	−1.02[a]	.25	−8.12[a]	.67	−2.56[a]	.29	−1.28[a]	.25
Cohort 2	−1.04[a]	.19	−2.90[a]	.50	−.40[a]	.15	−4.75[a]	.42	−1.11[a]	.23	−.90[a]	.22
Cohort 3	−.57[a]	.18	−2.31[a]	.42	.43[a]	.16	−3.35[a]	.39	−1.23[a]	.21	**−.47**	.22
Cohort 4	**−.26**	.18	**−.36**	.58	**.15**	.17	−2.52[a]	.56	**−.54**	.28	**.48**	.26
Cohort 5	**.20**	.22	**.61**	.47	.48[a]	.17	**−.93**	.47	**.18**	.26	.70[a]	.26
Cohort 6	.64[a]	.16	1.53[a]	.39	.97[a]	.16	**−.39**	.43	**.26**	.18	**.54**	.22

Note. Values in bold represent variables that do not show statistically significant change over 7 years.
[a] Mean deviations are significant at or beyond the .01 level of confidence.

structs of Inductive Reasoning, Spatial Orientation, Perceptual Speed, Numeric Facility, Verbal Comprehension, and Verbal Recall. Of course, this set of constructs does not represent a complete sampling of the domain of psychometric intelligence. Nor has each of the constructs been marked as broadly as some readers might prefer. It should be kept in mind that the test battery used in these analyses was assembled for a substantive longitudinal study of adult cognitive development (Schaie, 1983, 1994, 1996b). For the purposes of this study we elected to operate within the primary mental abilities framework, with abilities marked relatively narrowly, rather than considering a second order framework such as the Gc–Gf model (Horn, 1986), or conceptualizing intelligence in an even broader hierarchical framework (e.g., Sternberg & Berg, 1987).

In our earlier cross-sectional studies of factorial invariance (Schaie et al., 1989) we had demonstrated that if behavior is assessed across age with measures of satisfactory psychometric characteristics, such measures retain their conceptual position within the domain being studied and that the same number of factors suffice to describe that domain across groups differing in age. However, we also showed that the regression of the latent constructs upon the marker variables did not remain invariant across age groups and that the groups also differed significantly in their parameters for the factor variances and covariances. The positive inference from these findings was that it is realistic to compare the same basic constructs across groups widely differing in age in adulthood. The negative conclusion, however, was that the direct comparison of the individual observed tests that marked our constructs was questionable because of their differing regressions on the construct they are purported to measure.

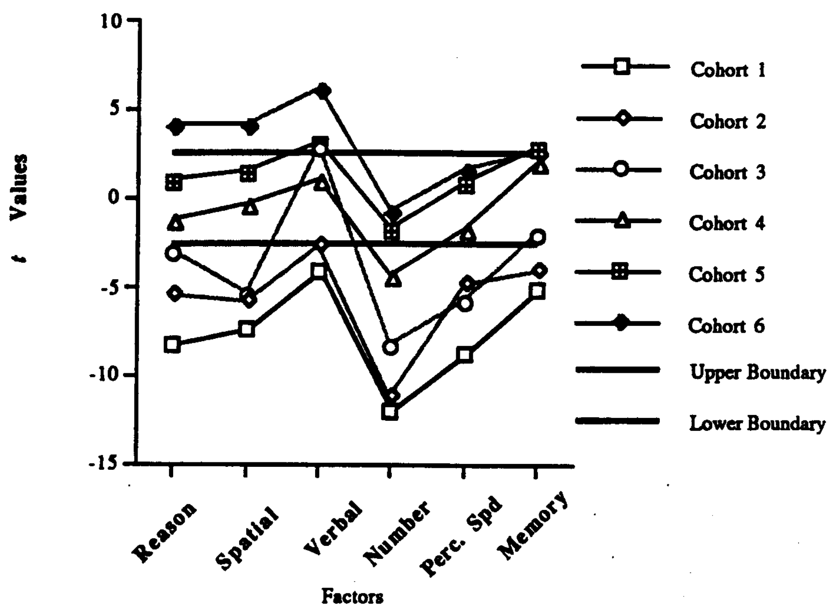

Figure 1. Values (*t*s) for changes in means across time within cohorts from 1984 to 1991.

Table 10
Mean Deviation Values and Standard Errors Between Groups

Group	Inductive Reasoning *M*	Inductive Reasoning *SE*	Spatial Orientation *M*	Spatial Orientation *SE*	Verbal Comprehension *M*	Verbal Comprehension *SE*	Numeric Facility *M*	Numeric Facility *SE*	Perceptual Speed *M*	Perceptual Speed *SE*	Verbal Recall *M*	Verbal Recall *SE*
1984 Assessment												
Cohort 1	−7.38[a]	.58	−15.29[a]	1.23	**.18**	.75	**−4.88**	1.90	−10.31[a]	.66	−4.19[a]	.44
Cohort 2	−4.56[a]	.55	−8.22[a]	1.06	**.90**	.70	**.49**	1.82	−6.14[a]	.58	−1.45[a]	.45
Cohort 3	−2.53[a]	.55	−3.52[a]	1.04	**.13**	.70	**−.59**	1.80	−3.19[a]	.56	−1.61[a]	.45
Cohort 4	**−.96**	.56	**−2.75**	1.13	**.01**	.72	**−1.03**	1.89	**−1.15**	.55	**−.89**	.45
Cohort 6	2.38[a]	.52	2.75[a]	.99	**−.95**	.73	**−1.02**	1.98	3.55[a]	.63	2.55[a]	.47
1991 Assessment												
Cohort 1	−9.45[a]	.58	−19.87[a]	1.25	**−1.21**	.77	−11.51[a]	1.97	−13.56[a]	.72	−6.29[a]	.48
Cohort 2	−5.81[a]	.56	−11.56[a]	1.06	**.07**	.70	**−4.23**	1.75	−7.55[a]	.63	−3.14[a]	.45
Cohort 3	−3.31[a]	.55	−6.36[a]	1.01	**.11**	.69	**−2.96**	1.73	−4.86[a]	.59	−2.85[a]	.45
Cohort 4	−1.44[a]	.56	−3.61[a]	1.12	**−.33**	.70	**−2.62**	1.78	−1.81[a]	.57	−1.18[a]	.44
Cohort 6	2.85[a]	.51	3.73[a]	1.01	**−.50**	.74	**−.94**	1.91	4.18[a]	.66	2.21[a]	.45

Note. Bolded mean deviation values did not differ significantly from the reference group. Means and standard errors for Cohort 5 are equal to zero.
[a] Statistically significant at or beyond the .01 level of confidence.

In this study we examined whether the same inferences would hold true when we tested the invariance of factor structures within the same group of individuals over time. The conclusions with respect to group differences continue to hold even when data for each subgroup are examined at two times of measurement. However, the within-group (longitudinal) data turn out to be far more robust. We can accept the weak factorial invariance model within all cohort groups. That is, for a sample ranging widely in age, we demonstrate the invariance of the factor loadings across a 7-year time interval at two times of measurement.

Given the sensitivity of LISREL to local disturbances, we reexamined the issues of cross-cohort (cross-sectional) invariance, to determine whether it was possible to demonstrate at least partial invariance across some even though not all cohorts. Our final accepted model indeed suggests that weak factorial invariance can be demonstrated across those cohorts now com-

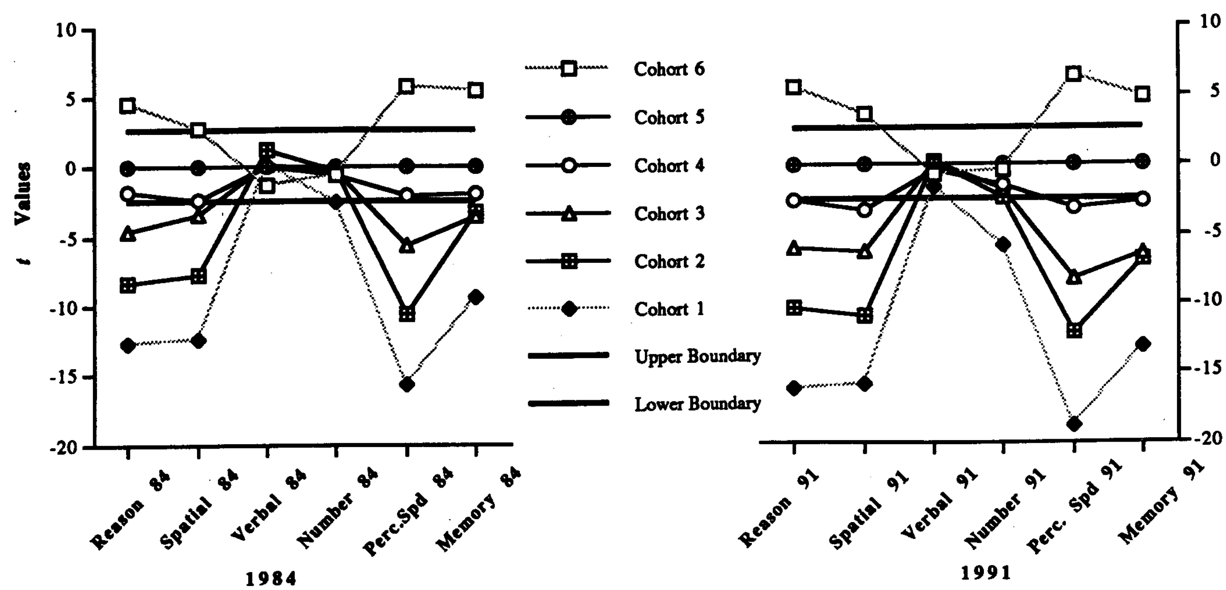

Figure 2. Values (*ts*) for differences in means between cohorts, from *t* Cohort 5 (second-youngest cohort).

posed of middle-aged and young-old participants. The invariance model had to be rejected, however, for young adults and old-old adults.

It is important also to point out that the discrepancies in factor loadings shown in young adulthood and advanced old age do not necessarily extend across the entire ability domain. Hence, our findings do not necessarily contradict those reported in other studies of equivalence of latent variables across age (cf. Horn & McArdle, 1992; Meredith, 1993). Indeed, we find invariance for the construct of Verbal Recall. For the other construct investigated, moreover, significant discrepancies in factor loading across groups are typically localized in some but not all markers of a given construct. Hence, investigators interested in age-comparative studies might benefit from selecting those marker variables for which superior invariance could be demonstrated.

We also addressed a number of substantive hypotheses with respect to changes in the factor space as well as changes and differences in factor means.

The data reported here shed some new light on the Wernerian (1948) concept of the differentiation and dedifferentiation of psychological domains in adulthood as extended to cognitive development by Reinert (1970). A more modern explanation for the hypothesized expansion and eventual contraction of the ability factor space may be found in shifts in the importance of processing speed as a major resource for effective cognitive behavior (cf. Lindenberger, Mayr, & Kliegl, 1993; Salthouse, 1994) or because of the ubiquitous role of sensory input as a final common pathway for behaviors in advance old age (Lindenberger & Baltes, 1994). Our data suggest that at least for much of adulthood the apparent contraction of the factor space may be an artifact of cross-sectional data. That is, we are able to accept the hypothesis of the invariance of the factor variance and covariances longitudinally at least over 7 years. Note, however, that for the oldest age group there is at least a trend for the covariances to increase across 7 years. On the other hand, we cannot accept invariance of the factor space cross-sectionally. That is, the shifts in the ability factor space, prior to the old-old stage of adulthood, may well be accounted for by demographic differences in successive cohort groupings. Nevertheless, we can once again only account for limited significant shifts in correlational magnitudes for those factor intercorrelations involving Spatial Orientation.

To examine age differences and longitudinal changes in factor means, we fit the estimated factor means to the accepted measurement model. The models, including the means structure, represent strong factorial invariance characteristics. Substantively, we found significant change over 7 years in factor mean scores for all of the abilities examined, except for Verbal Comprehension. However, the two youngest cohorts had significant gain or stability over 7 years. By age 60, significant within-group decrement was found only on Numerical Facility. Inductive Reasoning, Spatial Orientation, and Perceptual Speed showed initial decline by age 67, whereas Verbal Recall and Verbal Comprehension declined by age 74.

On the other hand, when we contrast the between-cohorts (cross-sectional) data, substantial differences are found on both test occasions for all latent variables except Verbal Comprehension and Numerical Facility. These findings are, of course, similar to those that we have previously reported for individual markers of the latent constructs (cf. Schaie, 1994, 1996b).

We conclude by suggesting, at least for the primary mental ability domain, that factorial invariance is demonstrable both cross-sectionally and longitudinally across middle adulthood. However, when young adults or old-old adults are included, robust invariance findings obtain only for longitudinal data. Cross-sectional studies including young adults and very old adults would therefore seem to require specific demonstrations of factorial invariance before the age-difference findings can be accepted.

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 raging research* (pp. 71–170). New York: Springer.

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., & Bonnet, 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: Wiley.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445–455.

Browne, M. W., & Cudeck, 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., 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.

Cohen, J. (1957). The factorial structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology, 21,* 283–290.

Cunningham, W. R. (1981). Ability factor structure differences in adulthood and old age. *Multivariate Behavioral Research, 16,* 3–22.

Cunningham, W. R. (1982). Factorial invariance: A methodological issue in the study of psychological development. *Experimental Aging Research, 8,* 61–65.

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

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging, 1,* 159–171.

Hertzog, C., & Schaie, K. W. (1988). Stability and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures. *Psychology and Aging, 3,* 122–130.

Horn, J. L. (1986). Intellectual ability concepts. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 35–78). Hillside, NJ: Erlbaum.

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 a the ethereal concept of factor invariance. *Southern Psychologist, 1,* 179–188.

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. (1989). *LISREL 8: A guide to the program and applications* (2nd ed.). Chicago: Scientific Software.

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age. *Psychology and Aging, 9,* 339–355.

Lindenberger, U., Mayr, U., & Kliegl, R. (1993). Speed and intelligence in old age. *Psychology and Aging, 8,* 207–220.

Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics, 19,* 201–215.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Meredith, W. (1964a). Notes on factorial invariance. *Psychometrika, 29,* 177–185.

Meredith, W. (1964b). Rotation to achieve factorial invariance. *Psychometrika, 29,* 187–206.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Nesselroade, J. R., & Labouvie, E. W. (1985). Experimental design in research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 35–60). New York: Van Nostrand Reinhold.

Reinert, G. (1970). Comparative factor analytic studies of intelligence through the human life-span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 468–485). New York: Academic Press.

Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology, 30,* 240–259.

Schaie, K. W. (1958). Rigidity-flexibility and intelligence: A cross-sectional study of the adult life-span from 20 to 70. *Psychological Monographs, 72*(462, Whole No. 9).

Schaie, K. W. (1977). Quasi-experimental designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 39–58). New York: Van Nostrand Reinhold.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.

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

Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainard (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 241–272). New York: Springer-Verlag.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49,* 304–313.

Schaie, K. W. (1996a). Intellectual development in adulthood. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 266–286). New York: Academic Press.

Schaie, K. W. (1996b). *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. (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., 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. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229–239.

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 semiannual meeting of the Psychometric Society, Iowa City, IA.

Sternberg, R. J., & Berg, C. (1987). What are theories of adult intellectual development theories of? In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 3–23). New York: Ablex.

Stricker, L. J., & Rock, D. A. (1987). Factor structure of the GRE general test in young and middle adulthood. *Developmental Psychology, 23,* 233–242.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

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

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

Werner, H. (1948). *Comparative psychology of mental development.* New York: International Universities Press.

White, N., & Cunningham, W. R. (1987). The age comparative construct validity of speeded cognitive factors. *Multivariate Behavioral Research, 22,* 249–265.

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

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

Willis, S. L., & Schaie, K. W. (1986). Training the elderly on the ability factors of spatial orientation and inductive reasoning. *Psychology and Aging, 1,* 239–247.

Zelinski, E. M., Gilewski, M. J., & Schaie, K. W. (1993). Three-year longitudinal memory assessment in older adults: Little change in performance. *Psychology and Aging, 8,* 176–186.

Received September 5, 1996
Revision received March 8, 1997
Accepted March 8, 1997 ■