Psychology and Aging
1988, Vol. 3, No. 2, 122–130

0882-7974/88/$00.75

Stability and Change in Adult Intelligence: 2. Simultaneous Analysis of Longitudinal Means and Covariance Structures

Christopher Hertzog
School of Psychology
Georgia Institute of Technology

K. Warner Schaie
Pennsylvania State University

We analyzed data on psychometric intelligence from the Seattle Longitudinal Study, simultaneously estimating longitudinal factors, their covariance structure, and their mean levels. Data on five Thurstone Primary Mental Abilities subtests were available for 412 adults, ages 22–70 at first test, who were tested three times at 7-year intervals. A previous longitudinal factor analysis had shown high stability of individual differences (covariance stability) in general intelligence for three adult age groups. We extended that model to estimate factor means. All three age groups showed high levels of covariance stability, but differed sharply in their mean profiles. The young group showed increasing levels of general intelligence, the middle-aged group had stable levels of intelligence, and the old group showed salient, approximately linear, decline. The patterns of stability in middle-age, followed by mean decline and high covariance stability in old age, suggest a normative developmental transition from a stability pattern to a decline pattern of general intelligence, with the inflection point occuring somewhere around age 60.

An important issue in the study of adult intellectual development concerns whether levels of intelligence remain stable with advancing age. There is general agreement that the average level of performance on certain psychometric measures of intelligence declines with age, although there is great debate as to (a) the ubiquity of decline, (b) the proper interpretation of decline in psychometric performance, when it occurs, and (c) the practical importance of the magnitude of age-related decline (e.g., Baltes, Dittman-Kohli, & Dixon, 1984; Botwinick, 1977; Dixon, Kramer, & Baltes, 1985; Horn, 1985; Horn & Donaldson, 1976, 1980; Schaie, 1983). At the center of the disagreements in the literature regarding aging and intelligence has been Schaie's longitudinal studies of aging and primary mental abilities (see Schaie, 1983). The debate between Horn, Schaie, and others (e.g., Baltes & Schaie, 1976; Horn & Donaldson, 1976) covered a large number of issues associated with Schaie's sequential design, psychometric tests, and alternate theories and interpetations of aging and intelligence. Subsequent work by Schaie and Hertzog (1983) re-examined the issues with new data from Schaie's sequential samples. Their cohort-sequential analyses identified clear cohort differences in certain psychometric tests and identified statistically significant changes in multiple psychometrically defined abilities. For all five subtests of Thurstone's Primary Mental Abilities (PMA; Thurstone, & Thurstone, 1949), declines in performance (whether measured by longitudinal or cross-sectional sequences) were negligible until after age 50. Declines that were observed after age 50 were small, but became increasingly large after mean age 60. A somewhat surprising result, given earlier cross-sequential results from Schaie's data, was that the longitudinal sequences suggested decline after mean age 60 in all PMA subtests, although the decline began later for the PMA subtest Verbal Meaning (a test of recognition vocabulary). Schaie and Hertzog (1983) argued that these results required some minor modification of previous positions regarding the age of onset of intellectual decline, but that they supported the major conclusions of (a) age-confounded cohort differences in cross-sectional studies, (b) relative stability of mean performance levels into the 50s, with substantial declines *only* after age 60, and (c) some differences across subtests in the onset and magnitude of age-related performance declines (see also Dixon et al., 1985).

Although most of the gerontological literature has focused on the issue of stability of mean levels of intelligence with aging, *mean stability* is but one type of stability that can be assessed in longitudinal data. Another important type of stability is *stability of individual differences* (e.g., Baltes, Reese, & Nesselroade, 1977; Kagan, 1980; Schaie & Hertzog, 1985). This stability reflects the degree to which individuals differ in their developmental patterns of change (Baltes et al., 1977; Nesselroade & Labouvie, 1985; Schaie & Hertzog, 1985). Whereas stability of means is reflected in equivalent mean values at different developmental times, stability of individual differences is reflected in the covariance of a variable with itself over two points in time (see Baltes et al. 1977). In this article, we refer to stability of individual differences as *covariance stability* (see Hertzog & Nesselroade, 1987).

We report data collected as part of the Seattle Longitudinal Study, which has been supported over an extended period of time by grants from the National Institutes of Health, the National Institute of Child and Human Development, and the National Institute on Aging. Currently, the study is supported by Public Health Service Grant R01-AG04770 from the National Institute on Aging. The first author's effort on this project was also supported by Research Career Development Award K04-AG00335 from the National Institute on Aging.

The cooperation and support from members and staff of the Group Health Cooperative of Puget Sound is gratefully acknowledged.

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332-0170.

In a previous article, Hertzog and Schaie (1986) demonstrated that there is substantial covariance stability in intelligence across the adult life span. Hertzog and Schaie (1986) used a longitudinal factor analysis of data from the Seattle Longitudinal Study (SLS; Schaie, 1983) to show (a) that a general intelligence factor, *g*, could be identified for three age groups (young, middle-aged, and old), (b) that this *g* factor was defined equivalently by the PMA subtests in each age group and showed invariant factor loadings across longitudinal occasions, (c) that the covariance stability of *g* was high in all age groups, with longitudinal correlations of *g* with itself at or above .9 between successive longitudinal occasions, even in the older group, and (d) that there was substantial covariance stability in the five primary ability subtests, independent of *g*, as reflected in the proportion of variance in the PMA subtests determined by "test-specific" factors.

Hertzog and Schaie's (1986) results support the hypothesis that age changes in *g* are relatively consistent for same-aged individuals. Although there are individual differences in change patterns, these differences produce shifts in relative ordering of individuals that are small relative to the overall population variance in *g*. It is interesting that covariance stability was high in age ranges in which Schaie and Hertzog (1983) detected decline in the individual PMA subtests—namely, after age 60. This finding suggests only modest individual differences in the magnitudes of late-life decline in *g*.

We report a series of additional analyses designed to examine explicitly the mean level stability of *g* and, simultaneously, to estimate stability of individual differences in *g*. The results of these analyses demonstrate the independence of these two type of stability in the domain of psychometric intelligence. The analyses also were used to examine the question of inflection point for shifts from stability to decline in general intelligence.

The simultaneous examination of mean and covariance stability in longitudinal data is made possible by use of structural equation models to analyze means of latent variables (e.g., McArdle & McDonald, 1984; Sörbom, 1982). The longitudinal factor analyses reported by Hertzog and Schaie (1986) constitute an important precursor to simultaneous analysis of mean and covariance structures. Hertzog and Schaie found metric invariance in the *g* factor loadings between groups and across longitudinal occasions of measurement. Metric invariance is defined as equivalence in the unstandardized regression weights of variables on factors (see Horn, McArdle, & Mason, 1984). As discussed by several developmental methodologists (e.g., Baltes & Nesselroade, 1973; Labouvie, 1980a, 1980b; Schaie & Hertzog, 1985), an assumption of metric invariance is essential for allowing unambiguous interpretation of quantitative differences in mean levels of factor scores. The demonstration of metric invariance in *g* ensures that *g* is measured in equivalent units of measurement, so that differences in *g* factor means are uncontaminated reflections of mean level differences in the latent variable (see Labouvie, 1980a, 1980b; Schaie & Hertzog, 1985, for further discussion of this issue).

Given evidence of metric invariance, the simultaneous analysis of means and covariance structures requires introduction of the means into the structural equations of the longitudinal factor model already used by Hertzog and Schaie (1986). The critical questions of interest were (a) What is the magnitude of mean age changes in *g* at the different age levels studied? (b) Do age differences and age changes in *g* fully account for the mean changes in PMA subtests, or must different developmental trends of PMA means be modeled to account fully for the information in the means? and (c) Is there evidence for independence of stability of *g* means from the covariance stability of *g*?

Method

Subjects

The subjects in this study were participants in the Seattle Longitudinal Study conducted by Schaie and his associates (Schaie, 1983). The population consisted of members of a health maintenance organization (HMO) in the greater Seattle area. The population was defined as all of the members of the HMO as of 1956, the initial year of the longitudinal study, in order to minimize the probability of selection differences over time. All of the participants were unpaid volunteers who answered questionnaires and took part in a single psychometric test session. The participants, adults between the ages of 20 and 74 years at the first test, represented a range of socioeconomic and ethnic groups (although the population defined by the HMO membership in 1956 was predominantly White and somewhat more affluent than the general Seattle population). Further details on the population and sampling procedures may be found in Schaie (1983).

Sequential Sampling Design

The longitudinal samples studied here are a subset of the sequential samples collected in the SLS. The sampling plan of the SLS is discussed more fully in Schaie (1983), and the present sample is defined explicitly in Hertzog and Schaie (1986). Briefly, we restrict our analysis here to two 14-year longitudinal samples (first tested in 1956 or in 1963). Data from the two longitudinal sequences were partitioned into a hybrid sequential data matrix described in Table 1. The partitioned data matrix forms three age groups for simultaneous analysis.

Variables

As part of a larger psychometric battery, all of the subjects were administered the 1948 version of the SRA Primary Mental Abilities Test, Form AM 11-17 (Thurstone & Thurstone, 1949). The 1948 PMA includes five subtests, all of which are timed and have significant speed components in adult samples (see Schaie & Hertzog, 1983): (a) Verbal Meaning—a test of recognition vocabulary, (b) Space—a test of spatial relations requiring mental rotation of figures in a two-dimensional plane, (c) Reasoning—a test of inductive reasoning requiring recognition and extrapolation of patterns of letter sequences, (d) Number—a test of the ability to solve simple two-column addition problems quickly and accurately, and (e) Word Fluency—a test of the ability to retrieve words from semantic memory according to an arbitrary syntactic rule (words beginning with the letter *s*). Scoring followed the PMA manual: Verbal Meaning and Reasoning were scored in terms of the number of correct items, Space and Number were scored by subtracting incorrect items (comission errors) from the total number of correct items, and Word Fluency was scored by tallying the number of unique, admissable words generated during the allotted time.

Models and Statistical Procedures

The longitudinal factor model used is an application of a generic longitudinal model described in some detail by Jöreskog and Sörbom (1977; see also Hertzog, in press; Horn & McArdle, 1980; Schaie & Hertzog, 1985). A detailed description of the model may be found in

Table 1
Reparameterized Sequential Sample for Multiple Group Analysis

Group/sample	Cohort (*M* birth year)	Age: Occasion 1	Occasion 2	Occasion 3	*n*
Group 1					
1	1931	25	32	39	21
1	1924	32	39	46	26
2	1938	25	32	39	22
2	1931	32	39	46	40
M		30	37	44	
Total					109
Group 2					
1	1917	39	46	53	27
1	1910	46	53	60	32
2	1924	39	46	53	51
2	1917	46	53	60	50
M		42	49	56	
Total					160
Group 3					
1	1903	53	60	67	28
1	1896	60	67	74	15
1	1889	67	74	81	13
2	1910	53	60	67	48
2	1903	60	67	74	18
2	1896	67	74	81	21
M		58	65	72	
Total					143

Hertzog and Schaie (1986). The model specified an occasion-specific *g* factor at each longitudinal occasion. The factor covariance matrix modeled the variances and covariances of *g* at the different occasions of measurement, and the residuals in the PMA subtests were modeled as having test-specific covariances (e.g., the residuals for Verbal Meaning were allowed to covary across longitudinal occasions). The specification of longitudinal models including factor means is relatively complex (Jöreskog & Sörbom, 1984; McArdle & Epstein, 1987; Sörbom, 1982). The critical features are (a) a vector of location constants, analogous to grand means, (b) representation of latent variable means as regressions on a fixed constant and modeled in the LISREL GAMMA parameter matrix, and (c) the assumption that the means of all residuals are zero in the population. The vector of location constants identifies an intercept for each observed variable (PMA subtest). In longitudinal analysis of multiple groups, these location parameters are constrained equal both across longitudinal occasions and between the multiple age groups. Given data containing neither group differences nor longitudinal changes in means, this location parameter vector would perfectly account for the mean structure. Thus, the model with factor means will be meaningful only if there are either group differences or longitudinal changes in observed variable means that the model may attempt to structure as a function of the factor means.

Identification of the location parameters and the factor means is achieved by fixing the mean of *g* to zero for one age group at one longitudinal occasion. In the models reported, we fixed the *g* mean for the middle-aged group at the first occasion (mean age 42) at zero. This procedure then enables the remaining factor means to be estimated as deviations from this reference point (see Jöreskog & Sörbom, 1984; Sörbom, 1982) for additional details. The fact that factor means are modeled as regression of factors (i.e., *g*) on a constant requires the assumption that the means of the residuals are zero. This is an unlikely assumption, given that we expect age trends in mean levels to vary across PMA subtests (independent of their relation to *g*). It is, however, possible to estimate residual component means by moving these parameters into the latent variable vector in LISREL.[1]

All of the models were estimated in either LISREL V or VI (Jöreskog & Sörbom, 1984) using maximum likelihood estimation. In structural modeling, model fit can be assessed by likelihood ratio chi-square, as well as relative fit indices provided by the program. These indices are of less value in models with means, however, so we report a decomposition of overall model fit into (a) fit of the covariance structure model and (b) fit of the mean structure model (see Bentler & Bonett, 1980; Sobel & Bohrnstedt, 1985). The relative fit index for the means may be interpreted as an index of the proportion of information in the mean structure, adjusted for location parameters, accounted for by the model.

The procedures used here are unabashedly exploratory in nature. The goal is to use the LISREL model to explore descriptive developmental hypotheses about the longitudinal mean and covariance structures of the PMA subtests. This use of a generic longitudinal factor model is an appropriate application of structural equation techniques, which are ideal for exploratory multivariate modeling of longitudinal data (Hertzog, in press; McArdle & Epstein, 1987). This study cannot and should not be considered to represent a confirmatory analysis, in the philosophical sense of the term.

Results

The first model we estimated fixed the *g* factor means at zero in all three age groups, but allowed all location parameters to be freely estimated. This model fits the 15 means of each age group with 15 freely estimated location parameters. There is

[1] A listing of the LISREL VI specifications for models with factor and residual means is available from the first author.

Table 2
Goodness-of-Fit for Longitudinal Factor Model With Means

Model	χ^2	*df*	F^a	*p*
M_s (saturated)	287.68	248	.352	.048
M_n (null in means)	642.02	288	.785	.000
M_1 (*g* factor means)	467.59	280	.572	.000
M_2 (*g* factor means; all 0 in middle-aged)	470.88	282	.575	.000
M_3 (*g* and test-specific factor means)	338.76	270	.414	.003
M_4 (*g* and residual means for V, S, N, W)	299.05	254	.366	.027

Note. V = Verbal Meaning; S = Space; R = Reasoning; N = Number; W = Word Fluency.
[a] LISREL fitting function at minimum.

a one-to-one correspondence between location parameters and sample means, and as such, the location parameters are just-identified. This model is therefore saturated with respect to the means, using Bentler and Bonett's (1980) definition. The fit of the model, denoted M_S, is reported in Tables 2 and 3. As expected, this model fit the same as the model ignoring means reported by Hertzog and Schaie (1986), and yielded an identical longitudinal factor solution. A second preliminary model, following recommendations of Bentler and Bonett (1980), was a null model in the means. This model specified five location parameters, one for each PMA subtest, and constrained these parameters to fit the means of all three longitudinal occasions for all three age groups. Thus, the 45 population means were fit with five location parameters. This null model, M_n, would have a fit equal to the saturated model, M_S, if there were no group differences or longitudinal changes in PMA subtest means to structure as part of the analysis. There was, however, a substantial, statistically significant difference between the two models, as seen in the first model comparison reported in Table 3. Clearly, there was longitudinal and age group variation in the PMA means, and the task of the analysis was to structure this variation in terms of the longitudinal factor model.

The first substantive model of interest specified *g* factor means in all three age groups. Interpretation of the fit of these substantive models must be made on the basis of relative differences from the null and saturated models, so that one can evaluate fit to the means ignoring (assuming) the basis specification and fit of the longitudinal factor model (Bentler & Bonett, 1980; Sobel & Bohrnstedt, 1985). In essence, the difference between the null and saturated models defines a range of possible fits of models structuring means in the longitudinal analysis. The critical question is how close a model with structured means comes to the fit of the model that is saturated in the means (or conversely, how far it has come from the poor fit of the null model).

As shown in Table 3, this first substantive model, M_1, improved meaningfully on the fit of the null model, although there was still a significant difference between M_1 and M_S. The relative fit of the new model is best indexed by the Sobel and Bohrnstedt (1985) relative fit index, denoted as δ in Table 3. The fit of .49 indicates that about half of the variation in the means had successfully been structured by M_1.

One interesting outcome of model M_1 was that the *g* factor means for the middle-aged adults were not significantly different from zero, relative to their standard errors. In models of this type, these estimated factor means are scaled as deviations from the fixed zero mean (age 42 for the middle-aged population). Therefore, the finding of essentially zero *g* means at ages 49 and 56 for the middle-aged group indicated no statistically significant change in mean level of *g* over this age range. A second model, M_2, incorporated this feature by fixing the *g* means to zero for all three ages of the middle-aged group. This model did not fit more poorly than M_1.

The fact that M_2 fit significantly worse than M_S implied that the assumption of no mean variation in the residuals for the PMA factors had to be abandoned. That is, it was not possible to model age-group differences and age changes in PMA means solely as a function of age differences and age changes in *g* factor means. Apparently, the primary abilities measured by the PMA have variations in the means that are saliently different from the behavior of the *g* factor means.

A logical possibility is that there are age group differences in subtest-specific means, but no age group differences in patterns

Table 3
Comparisons of Fit Between Alternative Models With Factor Means

	M_n		M_s				Comparison		
Model	$\Delta\chi^{2a}$	Δdf	$\Delta\chi^{2b}$	Δdf	δ^c	Comparison	$\Delta\chi^2$	Δdf	$\Delta\delta^d$
M_s	—	—	—	—	—		—	—	—
M_n	—	—	—	—	—	M_n–M_s	354.34	40	—
M_1	174.43	8	179.91	32	.492	—	—	—	—
M_2	171.94	6	182.40	34	.485	M_1–M_2	2.49	4	.007
M_3	303.26	18	51.08	22	.857	M_2–M_3	128.83	10	.365
M_4	342.97	34	11.37	6	.968	M_1–M_4	168.54	28	.483

[a] Difference in χ^2 between model and M_n (null model).
[b] Difference in χ^2 between model and M_s (saturated model).
[c] Relative fit index for fit to the mean structure.
[d] Change in relative fit index in means for models under comparison.

of age changes in the primary ability means. Such a pattern could arise if age changes in the primary abilities were solely a function of age changes in *g*, but there were also differential patterns of cohort effects across the primary ability means. Our previous work (Hertzog & Schaie, 1986), modeling both *g* and PMA test-specific factors, provided a convenient means of testing this hypothesis. We used a model that specified eight factors in each age group: (a) three *g* factors, one at each longitudinal occasion, and (b) five test-specific factors, one for each PMA subtest. We estimated factor means for all eight factors, achieving identification of the test-specific factor means by fixing all five test-specific factor means for the middle-aged group to zero. This model, M_3, allowed the *g* factor means at ages 49 and 56 to be freely estimated in the middle-aged group, as in model M_1. We did not wish to assume mean stability in *g*, even though that was suggested from the M_2–M_1 comparison. It could have been the case that the stable *g* factor means in the middle-aged group in the previous models were an artifact of model misspecification.

Model M_3 also constrained the test-specific factor loadings to be equal over the three age groups (see Hertzog & Schaie, 1986). The equality constraints on test-specific factor loadings did not permit any of the age-group differences in mean changes to be modeled by the test-specific factor means. Group differences in mean change on the PMA variables could only be reflected in the *g* factor means.

Table 2 reports the fit of M_3. The model fit significantly better than M_1, indicating there were statistically significant age group differences in test-specific factor means. However, the model still did not approximate the fit of M_S, requiring rejection of Model M_3. It was also still the case that the *g* factor means did not differ significantly between ages 42 and 56 for the middle-aged group. We concluded that there were age-group differences in PMA subtest means, but that there are also differential age changes for the PMA subtest means, independent of *g*. We also concluded that it was still plausible to maintain the assumption of no age changes in *g* in the middle-aged group.

We next proceeded by fitting a series of models allowing residual means. This approach was needed to allow for age-group differences in patterns of mean age changes on the primary abilities. This series of models proceeded in exploratory fashion. Large mean residuals (differences between sample means for the PMA subtests and PMA means predicted from the model parameters) and salient LISREL modification indices were used to indicate a need for structuring additional mean parameters. Unlike M_3, these models specified a separate PMA residual "factor" at each longitudinal occasion, permitting both *g* and the PMA residuals from *g* to display age-related change. After a series of model modifications, we arrived at a model that did not differ significantly from the saturated model. This model allowed residual means for Word Fluency, Number, Verbal Meaning, and Space. This modified model, M_4 in Table 2, achieved a relative fit index of .97 to the means, indicating excellent fit. Of course, this fit was achieved by adjusting to the sample means, and can therefore be treated only as a descriptive index of the success of the model modification process.

One of the major reasons for fitting additional models to the means was to ensure that the estimated age changes and age differences in *g* means were not inappropriately biased by the incorrect assumption of no residual means. Hertzog and Carter (1982) previously demonstrated that group differences in intelligence factor means were affected by the specification error of zero residual means. Table 4 reports the *g* factor means for the four substantive models, M_1 through M_4. Irrespective of the model, the relative pattern of *g* factor means in the three age groups remained the same. The *g* factor means increased from mean age 30 to mean age 37 in the young group, and then remained relatively stable through age 44. The *g* factor exhibited mean stability from mean age 42 through mean age 56 in the middle-aged group. Finally, *g* showed substantial decline from mean age 58 through mean age 72 in the old group. The mean decline in *g* in the old group was roughly linear over the 14-year period. The comparable pattern of *g* mean behavior is particularly important in Model M_4, in which it was most likely that the apparent age changes in *g* estimated in Models M_1 through M_3 would change as a function of specifying longitudinal changes in the PMA residuals as well. The fact that conclusions regarding the behavior of *g* means were not altered by specifying longitudinal variation in PMA residual means indicated that the mean patterns were unlikely to be an artifact of model specification.

Approximate 99% confidence intervals around the factor means can be calculated by subtracting and adding 2.5 *SE*s to the estimated *g* factor means. Inspection of Table 4 clearly showed that these 99% confidence intervals did not include zero for any of the freely estimated means in the old and young groups. As these means are deviation contrasts from the middle-aged *g* means, we concluded there were reliable age group differences in means. The significant differences included comparisons between the different groups at roughly comparable ages. That is, the young group at age 44 (Occasion 3) differed significantly from the middle-aged group at age 42 (Occasion 1), as did the middle-aged group at mean age 56 (Occasion 3) from the old group at mean age 58 (Occasion 1). Although the hybrid sequential design does not completely unconfound age changes and cohort differences, it seems likely that these differences reflect cohort differences in the mean levels of *g*.

Table 5 reports the residual means estimated in the final model, M_4. These means must be interpreted with care. They represent mean patterns in the PMA subtests orthogonal to the trends mediated through *g*. The first feature of note involves the residual means for Word Fluency and Number in the middle-aged group. Although the *g* means showed no age-related changes in the middle-aged, the residuals for Word Fluency and Number did change. There were small but statistically significant declines in Word Fluency and Number between mean ages 42 and 56. There is a second noteworthy feature of the residual means in Table 4. It seems that the large age-group (cohort) differences in *g overestimated* age group differences in Number and Verbal Meaning. This was shown by the large negative means in the young group for these two PMA subtests, as well as the large positive means for Number for the old group. Finally, there appeared to be modest levels of decline in Space for the old group (between mean ages 58 and 65) that was greater than the decline in Space predicted by *g*.

We do not report here the other parameter estimates from the longitudinal solution (e.g., factor covariances, factor loadings) because they differed trivially from the solution without means

Table 4
The g Factor Means for Alternative Longitudinal Models

		Model							
		M_1		M_2		M_3		M_4	
Group	*M* age	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Young									
g_1	30	1.61	0.60	1.62	0.59	8.54	3.26	2.82	0.65
g_2	37	2.76	0.57	2.78	0.57	10.11	3.49	3.99	0.65
g_3	44	2.70	0.56	2.71	0.55	9.87	3.39	3.50	0.62
Middle-aged									
g_1	42	0*	—	0*	—	0*	—	0*	—
g_2	49	0.10	0.17	0*	—	0.14	0.16	0*	—
g_3	56	−0.20	0.18	0*	—	−0.20	0.17	0*	—
Old									
g_1	58	−3.96	0.61	−3.97	0.60	−10.96	4.48	−4.20	0.64
g_2	65	−4.61	0.61	−4.62	0.61	−12.41	4.64	−4.78	0.64
g_3	72	−6.55	0.65	−6.57	0.64	−13.28	4.24	−6.22	0.66

Note. Asterisks denote fixed factor means. The *g* factor subscripts denote longitudinal occasion.

reported by Hertzog and Schaie (1986). However, one question remained regarding the factor covariance matrix for *g*. As reported in Hertzog and Schaie, there was an age-related increase in *g* factor variance in the old group. The old group also had greater overall variance in *g* than did the middle-aged and young groups. One possible explanation of these differences is that they are methodological artifacts. The old group was formed by pooling over a larger age span in order to achieve acceptable sample size for structural analysis (refer back to Table 1). In the present context, it was possible that the developmental changes in *g* factor means would differ if the youngest age group (mean age 53 at Occasion 1; age range, 50 to 56) were omitted from the analysis. To address this question, we redefined the old group to include only the individuals age 57 and older at first test, and re-ran the longitudinal model with this subsample. Briefly, this analysis showed (a) similar age declines in *g* means, but of greater magnitude, (b) higher variability in *g* in the old group, but (c) more homogeneity of *g* variance across the three longitudinal occasions. Thus, it appears that the increasing variability in *g* over time, found in the full sample, reflected differences in developmental patterns from ages 50 to 65, as opposed to heterogeneity of developmental trajectories of same-aged individuals in the latter part of the adult life span. The analysis thus provides further support for the argument of an inflection point around age 60, at which age decrements in PMA performance begin to accelerate. The increased variability in *g* in the older group is not, however, merely a methodological artifact of age-group definition.

Table 5
Residual Means in Final Model (M_4)

	Age group					
	Young		Middle-aged		Old	
Variable/ Occasion	*M*	*SE*	*M*	*SE*	*M*	*SE*
Verbal Meaning						
1	−5.10	1.01	0*		0.26	0.98
2	−4.75	1.07	0*		1.09	1.05
3	−3.65	1.03	0*		−0.49	1.08
Space						
1	0.58	1.15	0*		−1.19	1.01
2	0.98	1.22	0*		−2.68	1.01
3	1.76	1.20	0*		−2.56	1.03
Reasoning						
1	0*		0*		0*	
2	0*		0*		0*	
3	0*		0*		0*	
Number						
1	−5.56	1.32	0*		3.71	1.23
2	−5.58	1.40	0.28	0.44	5.12	1.28
3	−6.03	1.31	−1.62	0.43	3.38	1.27
Word Fluency						
1	−1.45	1.48	0*		4.98	1.45
2	−3.56	1.59	−1.43	0.68	2.77	1.46
3	−1.18	1.60	−2.08	0.69	2.36	1.49

Note. Asterisks denote fixed 0 parameters.

Discussion

The results from this analysis amplify and accentuate several issues regarding age changes in psychometric intelligence. First, the results extend Schaie's (1983) work on age patterns in multiple primary intellectual abilities to the level of general intelligence, as measured by the *g* factor defined from the PMA subtests. We found a pattern of age changes in *g* factor means highly consistent with previous univariate results (e.g., Schaie & Hertzog, 1983). There were small increases in *g* in early adulthood (through mean age 32), stability in *g* means through middle age (until mean age 56), and substantial decline in late life. We explicitly tested the hypothesis that there was no decline in *g* in the middle-aged group at two different junctures, and could not reject the hypothesis. Moreover, the age changes that were estimated as part of this hypothesis test were so small as to be trivial in importance. On the other hand, we did find evidence of some decline in the middle-aged group on the PMA subtests Word Fluency and Number, independent of *g*.

The results also suggest substantial cohort differences in *g* means. The age groups differed not only in terms of mean age at initial test but also in birth cohort membership. The fact that the middle-aged group at mean age 56 performed significantly better on *g* than did the old group at mean age 58 surely indicates salient cohort differences in these data, as already detailed by Schaie (1983).

The unique contribution of this study, in terms of estimating age changes in PMA means, stems from the fact that the mean differences are estimated at the level of the *g* factor. Because these estimates are based on the simultaneously estimated factor pattern weights, they represent optimal estimates of *g* factor means that are not contaminated by mean patterns specific to the primary abilities themselves. Moreover, the analysis permitted the evaluation of mean trends in the primary abilities after they have been residualized with respect to *g*.

An additional contribution of the present analysis is that it permits independent evaluation of mean stability and covariance stability in *g*. These results demonstrate concretely the independence of these two types of stability. In all three age groups, individual differences in *g* were highly stable over the 14-year period. Yet each age group showed dramatically different age trends in *g*. In the young group, *g* increased to a stable plateau. In the middle-aged group, *g* means remained stable, but in the old group, substantial *g* decline was observed.

The change in mean patterns across the age groups, coupled with the high degree of covariance stability across the life span, has important implications for several prominent hypotheses about adult intellectual development. It is often the case, especially recently, that *g* is identified with basic intelligence (e.g., Jensen, 1982). Given (a) the widely accepted notion that there is multidirectionality in age trends in ability, such that some, but not all, abilities show age-related declines (e.g., Baltes et al., 1984; Botwinick, 1977; Horn & Donaldson, 1980) and (b) the accepted argument that it is measures of fluid intelligence (Horn, 1985; Horn & Donaldson, 1980), or alternatively, Wechsler-type performance tests (Botwinick, 1977; Salthouse, 1982) that manifest early decline, one would expect that *g*, as measured here, would be the prime candidate for evidencing decline from ages 25 to 55. To the contrary, it appears to be the case that *g* manifests *both* mean stability and covariance stability in middle age in the Seattle Longitudinal Sample.

How can this discrepancy be explained? One possible explanation is that the *g* factor estimated by the PMA variables is highly specific to the variables or to the samples, and hence is in some way a poor measure of the construct of general intelligence. This possibility seems relatively implausible. The *g* factor loadings estimated here are highly consistent with those found by Thurstone and Thurstone (1941) for these tests, and show a pattern of loadings consistent with a plethora of studies from the psychometric literature. The best indicator of *g* in the PMA, judged from our factor loadings, is Reasoning. This subtest, a measure of induction, is probably the best indicator of general intelligence and of the Horn-Cattell second-order fluid intelligence factor in the PMA (Horn & Donaldson, 1976). Not only did the Reasoning test load highly on *g*, but the Reasoning means in all age groups were well fit by the models specifying no age-related changes in *g* in the middle-aged group. Although we have estimated the single higher order *g* factor here, as opposed to fluid intelligence, Gustaffson (1984) recently reported hierarchical factor results from multiple intelligence tests that suggest that the *g* factor is isomorphic with fluid intelligence.

Thus, it would seem that the hypothesis of early decline in *g* is not supported by these data. The best model for the development of *g* in middle-age is a model of stability in both means and individual differences. One could argue that the generalizability of these results is limited because individuals who manifest early decline are more likely to drop out of longitudinal studies. Perhaps so, but the finding of mean stability of *g*, even in a select subpopulation, argues against the ubiquity of early age declines in *g*. There is evidence in these data of decline in two PMA subtests, Word Fluency and Number, in the middle-aged group. We suggest that, barring the sort of nonnormative events that lead to early mortality, individuals appear to maintain stable performance levels of *g* until sometime after age 50.

However, the developmental pattern of *g* begins to change dramatically between ages 50 and 60. After mean age 58, we found substantial, statistically significant decrements in mean levels of *g*. This decline was observed in an age group in which the covariance stability of *g* remained quite high. These results, then, offer little support to the hope that age-related decline in *g* is somehow nonnormative or is restricted to a small subpopulation of older individuals. We did find increased variance in *g* in the middle-aged and older groups, suggesting some small differences in developmental trajectories between those individuals in their 50s and those in their 60s. However, the longitudinal increases in *g* variance in the older group—crucial to the argument of different developmental trajectories in old age—were eliminated when the old group was restricted to individuals age 57 and older at first test.

The fact that it was necessary to fit residual mean factors, varying in age patterns, provides support for the arguments of Baltes and colleagues (e.g., Baltes et al. 1984) that intelligence is both multidimensional and multidirectional in its development. For example, the fact that young adults have *lower* means on the Verbal Meaning residuals suggests that the *g* factor means overestimate the age differences in vocabulary, even though Verbal Meaning has high loadings on *g*. This pattern is also observed for the Number and Word Fluency residual means, and may suggest reversed cohort differences on these tests when *g* is statistically removed from these tests. The pattern of Space residual means in the old group indicates greater decline between ages 58 and 65 on spatial ability than is predicted by *g*. Some caution is in order in interpreting these residual means. Our data only permit estimation of factor means for *g*. These residual means do not have the same status as means estimated in models with multiple measures of each primary ability, being much more likely to be specific to the PMA subtest than would primary ability factor means.

The analysis provides relatively little evidence of substantial individual differences in intraindividual change in general intelligence. To the contrary, these findings of differential age group patterns in *g* means, coupled with high degree of covariance stability in all age groups, suggest a relatively *normative* developmental transition in *g*. That is, it appears that most individuals make a transition from a stability to a decline pattern of *g* development at some point between age 55 and age 70, with individual differences in the age of onset of this transition.

It is important to note that these inferences are based on population parameters, and that there are some individuals who do not show salient decline even into old age (Schaie, 1983). There may be greater heterogeneity of change for the primary abilities, as opposed to *g* (see Hertzog & Schaie, 1986). Nevertheless, the results suggest that the heterogeneity of developmental trends in *g* during old age is small when measured against the population variance.

The high degree of covariance stability is a descriptive phenomenon and should not be assumed to demonstrate the validity of biological causes of age changes in *g*. Stability does not imply immutability, and Schaie and Willis (1986) have demonstrated significant training gains in inductive reasoning in individuals with prior histories of decline in this ability (all of whom were, in fact, part of the samples used in the present analysis).

In a sense, these results contradict aspects of the arguments made by both sides of the debate regarding the nature of intellectual decline manifested in the Seattle Longitudinal Study (Baltes & Schaie, 1976; Horn & Donaldson, 1976). The results appear, however, consistent with the updated perspectives of both Horn (1985) and Baltes and his colleagues (e.g., Baltes et al., 1984; Dixon et al., 1985). The key involves an assessment of the kinds of abilities measured in timed psychometric tests such as the Thurstone PMA, and hence, the nature of the *g* factor extracted from it. Evidence from a number of studies have shown that Thurstone-type tests of primary abilities have high correlations with speed of basic perceptual processes in adult samples (Cornelius, Willis, Nesselroade, & Baltes, 1983; Hertzog, 1987; Horn, Donaldson, & Engstrom, 1981). Schaie originally selected the adolescent form of the PMA for his study, and this form has limited item difficulty and substantial speed components in adult samples (e.g., Schaie, Rosenthal, & Perlman, 1953). The *g* factor estimated in this study was marked as highly by PMA Verbal Meaning as by PMA Reasoning. We have recently shown a strong relationship of PMA Verbal Meaning to a Perceptual Speed factor independent of its relationship to other vocabulary tests (e.g., ETS Advanced Vocabulary; Schaie, Willis, Hertzog, & Schulenberg, 1987). Thus, it appears that the PMA was constructed so as to maximize variance determined by what might be termed the *mechanics* of intelligence (e.g., Hunt, 1978), that is, the speed of basic cognitive processes needed for rapid decisions of low to moderate difficulty. Given that age-related slowing in information-processing speed is a highly normative developmental phenomenon (e.g., Birren, 1974; Salthouse, 1985), we can construct the following argument. The PMA manifests little age change in *g* prior to age 55 because *g*, as operationally defined by the PMA, emphasizes speeded solution of problems of limited difficulty. However, sometime after age 50, the age-related slowing in information-processing speed becomes a salient limiting factor in PMA performance, and *g* begins to decline dramatically. Individual differences in decline are minimized because (a) the PMA items are not optimally sensitive to the type of cognitive processes likely to maximize psychometric test performance in superior old adults (e.g., strategies for solving difficult problems, cognitive styles, and metacognitive processes; Baron, 1985; Dixon, in press; Sternberg, 1985) and (b) the ability domain covered by the tests is highly limited, excluding the types of abilities most likely to show increment and differential growth in adulthood, such as social cognition, domain-specific procedural knowledge, expertise, and postformal reasoning (Berg & Sternberg, 1985; Dixon et al., 1985; Labouvie-Vief, 1985; Rybash, Hoyer, & Roodin, 1986). Although important gains can be made by studying these other domains of cognition, we maintain that the study of cognitive mechanics, as they relate to performance on intelligence tests, remains a continuing priority for gerontology. A formal test of the cognitive mechanics interpretation of psychometric test performance in adulthood requires investigation of the nature of the information-processing skills tapped by Thurstone-type tests, research now ongoing in several laboratories.

References

Baltes, P. B., Dittman-Kohli, F., & Dixon, R. A. (1984). New perspectives on the development of intelligence in adulthood: Toward a dual-process conception and a model of selective optimization with compensation. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 6, pp. 34–76). New York: Academic Press.

Baltes, P. B., & Nesselroade, J. R. (1973). The developmental analysis of individual differences on multiple measures. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 219–252). New York: Academic Press.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods.* Monterey, CA: Brooks-Cole.

Baltes, P. B., & Schaie, K. W. (1976). On the plasticity of intelligence in old age: Where Horn and Donaldson fail. *American Psychologist, 31,* 720–725.

Baron, J. (1985). *Rationality and intelligence.* New York: Cambridge University Press.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Berg, C. A., & Sternberg, R. J. (1985). A triarchic theory of intellectual development during childhood. *Developmental Review, 5,* 334–370.

Birren, J. E. (1974). Translations in gerontology—from lab to life: Psychophysiology and the speed of response. *American Psychologist, 29,* 808–815.

Botwinick, J. (1977). Intellectual abilities. In J. E. Birren & K. W. Schaie, (Eds.), *Handbook of the psychology of aging* (pp. 580–605). New York: Van Nostrand Reinhold.

Cornelius, S. W., Willis, S. L., Nesselroade, J. R., & Baltes, P. B. (1983). Convergence between attention variables and factors of psychometric intelligence in older adults. *Intelligence, 7,* 253–269.

Dixon, R. A. (in press). Questionnaire research on metamemory and aging: Issues of structure and function. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and old age.* New York: Cambridge University Press.

Dixon, R. A., Kramer, D. A., & Baltes, P. B. (1985). Intelligence: Its life-span development. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement, and applications* (pp. 469–518). New York: Wiley.

Gustafsson, J. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179–203.

Hertzog, C. (1987). *The influence of cognitive slowing on age differences in intelligence.* Unpublished manuscript.

Hertzog, C. (in press). On the utility of structural regression models for developmental research. In P. B. Baltes, D. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 10). Hillsdale, NJ: Erlbaum.

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

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait–state distinction for the structural modeling of developmental change. *Child Development, 58,* 93–109.

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

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theory, measurements, and applications* (pp. 267–300). New York: Wiley.

Horn, J. L., & Donaldson, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31,* 701–719.

Horn, J. L., & Donaldson, G. (1980). Cognitive development in adulthood. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 445–529). Cambridge, MA: Harvard University Press.

Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research on Aging, 3,* 33–84.

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980's: Psychological issues* (pp. 503–541). Washington, DC: American Psychological Association.

Horn, J. L., McArdle, J. J., & Mason, R. (1984). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist, 1,* 179–188.

Hunt, E. (1978). The mechanics of verbal ability. *Psychological Review, 85,* 109–130.

Jensen, A. R. (1982). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Jöreskog, K. G., & Sörbom, D. (1977). Statistical models and methods for analyses of longitudinal data. In D. S. Aigner & A. S. Goldberger (Eds.), *Latent variables in socio-economic models* (pp. 285–325). Amsterdam: North Holland.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI user's guide.* Mooresville, IN: Scientific Software.

Kagan, J. (1980). Perspectives on continuity. In O. G. Brim, Jr. & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). Cambridge, MA: Harvard University Press.

Labouvie, E. W. (1980a). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980's: Psychological issues* (pp. 493–502). Washington, DC: American Psychological Association.

Labouvie, E. W. (1980b). Measurement of individual differences in intraindividual changes. *Psychological Bulletin, 88,* 54–59.

Labouvie-Vief, G. (1985). Intelligence and cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 500–530). New York: Van Nostrand Reinhold.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58,* 110–133.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 234–251.

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.

Rybash, J. M., Hoyer, W. J., & Roodin, P. A. (1986). *Adult cognition and aging: Developmental changes in processing, knowing, and thinking.* New York: Pergamon Press.

Salthouse, T. A. (1982). *Adult cognition: An experimental psychology of human aging.* New York: Springer.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam: North Holland.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A 21-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., & 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., Rosenthal, F., & Perlman, R. M. (1953). Differential mental deterioration on factorially "pure" functions in later maturity. *Journal of Gerontology, 8,* 191–196.

Schaie, K. W., & Willis, S. L. (1986). Can decline in intellectual functioning be reversed? *Developmental Psychology, 22,* 223–232.

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

Sörbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 183–195). North Holland: Amsterdam.

Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 152–178). San Francisco, CA: Jossey-Bass.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of intelligence.* New York: Cambridge University Press.

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence* (Psychometric Monographs, No. 2). Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1949). *Examiners manual: SRA Primary Mental Abilities Test* (Form 11–17). Chicago: Science Research Associates.

Received January 12, 1987
Revision received July 15, 1987
Accepted July 17, 1987 ■