Psychology and Aging
1986, Vol. 1, No. 2, 159–171

0882-7974/86/$00.75

Stability and Change in Adult Intelligence: 1. Analysis of Longitudinal Covariance Structures

Christopher Hertzog and K. Warner Schaie
Pennsylvania State University

We address two questions of central interest in adult intellectual development: the equivalence of psychometric tests' measurement properties at different ages, and the stability of individual differences in intelligence over time. We performed a series of longitudinal factor analyses using the LISREL program to model longitudinal data from Schaie's Seattle Longitudinal Study. The results indicate complete invariance in the loadings of five subtests of Thurstone's Primary Mental Abilities battery on a general intelligence factor. Individual differences in general intelligence were highly stable over 14-year epochs, with standardized factor correlations averaging about .9 between adjacent 7-year testing intervals. These results indicate that most individuals in this relatively select longitudinal sample maintained their relative ordering in intelligence.

One of the central questions in adult development regards the stability of adult intelligence—does intelligence decline with age, and if so, what is the magnitude of individual differences in patterns of change (e.g., Botwinick, 1977; Horn & Donaldson, 1980; Schaie, 1983)? The debate in the literature on the development of intelligence during adulthood has focused primarily on the stability of mean levels of intelligence—is there indeed decline, on average, on different intellectual abilities, and if so, what is the magnitude of such decline (e.g., Baltes & Schaie, 1976; Horn & Donaldson, 1976; Schaie & Hertzog, 1983)? The attention paid to stability of mean levels of intelligence has perhaps diverted the field from focusing on a different, critical—and in some senses more critical—type of stability: *stability of individual differences* in intelligence. How large are individual differences in magnitudes of age changes in intelligence during the adult years? Some developmental psychologists have suggested that adult development is characterized by increasing heterogeneity and by substantial individual differences in patterns of age change in intelligence and other cognitive capacities and skills (e.g., Baltes, Dittmann-Kohli, & Dixon, 1984; Hertzog, 1985; Schaie, 1983). Enhancement of optimal intellectual development through intervention (e.g., Schaie & Willis, 1986) requires as a first step the identification of differential patterns of aging and the isolation of the causes of such differences.

Measuring stability of individual differences in intelligence is somewhat more complex than measuring mean level stability. Although sequential sampling strategies using repeated, independent cross-sectional samples can be used to assess mean level stability (e.g., Schaie, 1977; Schaie & Hertzog, 1982), stability of individual differences can only be addressed by following individuals in a longitudinal panel design. Cross-sectional designs can only measure magnitudes of individual differences—as indicated by the variances—at a single point in time. At any given point in time, individual differences can be conceptualized as being determined by an earlier individual differences distribution and by subsequent individual differences in developmental change (see Baltes, Reese, & Nesselroade, 1977). Only a longitudinal design, by directly measuring change at the level of the individual, can be used to estimate the proportion of individual differences due to individual differences in change during preceding time periods (see Hertzog, 1985; Nesselroade & Labouvie, 1985; Schaie & Hertzog, 1985).

This study was designed to provide a careful and detailed examination of individual differences in intellectual change during adulthood. It also focuses on a second, critical issue identified by developmental methodologists regarding the assessment of change over time in variables such as intelligence. The issue is whether the constructs under study, and the measures of those constructs, are actually isomorphic at different ages. Can we assume that intelligence is the same construct at ages 25 and 75? Even if intelligence is unchanging, or continuous (Kagan, 1980) across the adult life span, is it the case that psychometric measures of intelligence are equally reliable and valid as measures of intelligence at different ages? Baltes and Nesselroade (1970) identified this issue as one of *measurement equivalence*—can we assume invariant measurement properties of empirical measures at different parts of the life span (see also Eckensberger, 1973)? As Baltes and Nesselroade indicated (see also Schaie, 1977; Schaie & Hertzog, 1985), the optimal method for assessing measurement equivalence is comparative factor analysis, in which the invari-

This article reports 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 for Child and Human Development, and the National Institute on Aging. Our work is currently supported by Grant R01-AG4770 from the National Institute on Aging.

Our thanks to William Meredith for advice and comments on our statistical models and results, and to an anonymous reviewer for helpful editorial suggestions. The cooperation and support from members and staff of the Group Health Cooperative of Project Sound is gratefully acknowledged.

Correspondence concerning this article should be addressed to Christopher Hertzog, who is now at the School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332, or to K. Warner Schaie at S-110 Human Development Building, The Pennsylvania State University, University Park, Pennsylvania 16802.

ance of the factor structure of the psychometric abilities is assessed. As discussed elsewhere (e.g., Cunningham, 1978; Schaie & Hertzog, 1982, 1985), the best approach to the invariance problem involves the use of confirmatory factor analytic methods to test the hypothesis of age-related invariance in the factor structure.

This is the first in a series of articles describing our use of covariance structures methods to analyze patterns of change and stability in adult intelligence using data from Schaie's Seattle Longitudinal Study (SLS). In this article we describe results from a longitudinal factor model that may be used to assess (a) the measurement equivalence of the Thurstone Primary Mental Abilities battery used in the SLS and (b) the extent to which individuals in the SLS vary in patterns of intellectual change during the adult years. The Primary Mental Abilities test was developed by Thurstone and Thurstone (1941, 1949) to measure factorially pure, but intercorrelated, intellectual abilities. Assessment of factorial invariance and stability of individuals with the Primary Mental Abilities is particularly relevant, given the influence of Thurstone's work on the field of psychometric intelligence. Our findings strongly support the measurement equivalence of the Thurstone battery across much of the adult life span. We also show that there is a surprising degree of stability of individual differences in intelligence in participants from the kind of long-term longitudinal sample obtained in the SLS.

Our conclusions are based on results from a set of relatively complex longitudinal covariance structures models of the type developed by Jöreskog and co-workers (e.g., Jöreskog & Sörbom, 1977). The longitudinal factor model developed by Jöreskog and others (Jöreskog, 1979; Jöreskog & Sörbom, 1977) may be viewed as a generalization of other longitudinal factor analysis (e.g., models by Corballis, 1973; Corballis & Traub, 1970). To set the stage for our report, we must first summarize the methodological features of these models and how their parameters may be used to assess stability and change in individual differences over time (see also Hertzog, in press; Horn & McArdle, 1980; Schaie & Hertzog, 1985).

Let us assume that an investigator has collected multiple measures of one or more latent variables in a longitudinal design. The measures may or may not be identical at each longitudinal measurement occasion, although in the SLS the same measures were collected at each time of measurement. The relations among these variables must be represented by the covariance matrix of the observed variables (a correlation matrix should *not* be analyzed; Jöreskog & Sörbom, 1977). Given this kind of replicated longitudinal design, confirmatory factor analysis may be used to specify and estimate a longitudinal factor model with the following features.

First, the same factor structure is hypothesized to exist at each longitudinal measurement occasion. This structure is represented in the *factor pattern* matrix, which contains the regression coefficients mapping variables on factors (factor loadings). In the analysis we report here, a general intelligence (*g*) factor was modeled at each longitudinal occasion. The factors thus specified in a longitudinal factor model are often termed *occasion-specific* factors.[1] In addition to the factor pattern matrix, the basic longitudinal model includes a *factor covariance* matrix, describing the relations among the factors within and between longitudinal occasions, and a *residual covariance* matrix. The primary parameters of interest are the factor loadings and the factor covariance matrix. The first step involves evaluation of the measurement equivalence of the observed variables. Measurement equivalence may be assessed by (a) evaluating the adequacy of the model postulating isomorphic occasion-specific factors (i.e., the same number of factors with the same configuration of factor loadings at each longitudinal occasion) and (b) determining the plausibility of a model constraining these factor loadings to be equal (invariant) over all longitudinal occasions. These factor loadings are raw-score (unstandardized) regression coefficients, and invariance of these coefficients (sometimes termed *metric invariance;* see Horn, McArdle, & Mason, 1984) implies unchanging relations of the observed variables to the factors (Meredith, 1964; Schaie & Hertzog, 1985). Procedures for assessing the fit of these models are described later in the article.

Given that the hypothesis of measurement equivalence is tenable, the second step in the longitudinal analysis shifts attention to the factor covariance matrix. The diagonal elements of this matrix—the factor variances—reflect the magnitude of individual differences at each longitudinal occasion. Changes in factor variances would therefore reflect changes in the overall magnitude of individual differences over time. The stability of individual differences across longitudinal occasions is reflected in the covariances of factors with themselves over time. If the covariance of a factor at Time 1 with itself at Time 2 is large and positive, then individuals are preserving their relative order about the factor mean between Times 1 and 2. On the other hand, a zero or near zero covariance would reflect a high degree of flux in individual differences between Times 1 and 2. As shown by Baltes, Reese, and Nesselroade (1977), a zero covariance would be consistent with large individual differences in the patterns of developmental change during that time period.

Given that the SLS is a sequential study, in which multiple longitudinal samples have been followed over time (see Schaie, 1979, 1983), it is possible to expand the longitudinal model to consider longitudinal changes in multiple age groups. The extension of the model to multiple group analysis has been described by Jöreskog and Sörbom (1980), and is relatively straightforward. The advantage of a multiple groups analysis in the present context is that it allows us to address the issue of age invariance in factor structure both longitudinally, within a group of individuals, and comparatively, across multiple age groups. The longitudinal samples we analyze include adults of a wide span of chronological ages who have been tested three times over a 14-year period. These multiple samples allow us to examine longitudinal invariance in factor structure over 14-year epochs, while also examining factorial invariance over the adult life-span by comparing the factor structures of multiple age groups.

Method

Subjects

The subjects in this study were participants in the Seattle Longitudinal Study conducted by Schaie and associates (Schaie, 1979, 1983). The population consisted of members of a health maintenance organization (HMO) in the greater Seattle, Washington, area. To minimize the prob-

---

[1] The model can be extended without difficulty to include different numbers of common factors at each longitudinal occasion, but that approach is unnecessary in our analysis.

ability of selection differences over time, the population was defined as all members of the organization as of 1956, the initial year of the longitudinal study. All participants were unpaid volunteers who answered questionnaires and took part in a psychometric testing conducted in a single session. The volunteers were recruited from a randomly drawn sampling frame of the HMO membership, stratified by age and gender. The participants were adults spanning the age range from 20 through 74, at first test, and representing a range of socioeconomic and ethnic groups. However, probability sampling was *not* employed, and the sample was therefore not necessarily representative of the entire HMO population. As was generally true of the Seattle population circa 1956, the sample is predominantly Caucasian and, reflecting the membership of the HMO, contains a higher proportion of middle- and upper-income individuals than did the total Seattle population. Further details on the population and sampling procedures may be found in Schaie (1979, 1983).

Sequential Sampling Design

The longitudinal samples studied here are a subset of the sequential samples collected in the SLS. Briefly, the design of the SLS consisted of repeated sampling from the population at 7-year intervals, beginning in 1956 and continuing through 1984. Each year of testing, a new cross-sectional sample was drawn from the population, and all previously tested individuals were contacted and recruited for participation in the longitudinal panel. Thus, each independent cross-sectional sample was transformed into a multiple-cohort longitudinal sequence (Baltes et al., 1977) by repeated testing of the same individuals. We restrict our analysis here to two 14-year longitudinal samples: Sample 1 consists of 162 subjects tested in 1956, 1963, and 1970, and Sample 2, 250 subjects tested in 1963, 1970, and 1977. The data from the two longitudinal sequences were partitioned into a hybrid sequential data matrix given in Table 1. This partition created three age groups (young, middle aged, and old) for simultaneous analysis. These age groups were formed under the assumption of no cohort differences in factor structure. Although it would have been desirable to test for both age-related and cohort-related measurement equivalence, sample sizes were insufficient for such purposes. Age-related changes in factor structure seemed more likely, a priori, and earlier work supported the assumption of no cohort differences in factor structure (Cunningham & Birren, 1980). As can be seen from Table 1, data from different birth cohorts were pooled to obtain the age groups.

Variables

As part of a larger psychometric battery, all of the subjects were administered the 1948 version of the SRA (Science Research Associates) Primary Mental Abilities (PMA) 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 (Schaie, Rosenthal, & Perlman, 1953). They are (a) Verbal Meaning—a test of recognition vocabulary, (b) Space—a test of spatial orientation requiring mental rotation 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. Scoring protocols followed the PMA manual: Verbal Meaning and Reasoning are scored in terms of the number of correct responses; Space and Number are scored by subtracting commission errors from the total number correct; and Word Fluency is scored by tallying the total of unique, admissible words generated.

Statistical Procedures

All of the models described were tested using the LISREL V program of Jöreskog and Sörbom (1981). The analyses reported in this article used only one of LISREL's two-factor analysis measurement models. In LISREL notation, the measurement model may be specified as

$$x = \Lambda\xi + \delta, \qquad (1)$$

which in matrix form specifies a q-order vector of observed variables, x, as a function of their regression on n latent variables (factors) in ξ, with regression residuals δ. The $q \times n$ matrix Λ contains the regression coefficients (factor loadings). Equation 1 implies that the covariance matrix of the observed variables in the populations, Σ, may be expressed as

$$\Sigma = \Lambda\Phi\Lambda' + \Theta, \qquad (2)$$

where Λ is as before, Φ is the covariance matrix of the ξ, and Θ is the covariance matrix of the δ. Equation 2 is a restricted factor analysis model that can be extended to multiple groups (Jöreskog, 1971).

The parameters of LISREL's restricted factor analysis model are estimated by the method of maximum likelihood, provided that a unique solution to the parameters has been defined by placing a sufficient number of restrictions on the equations in Equation 2 to identify the remaining unknowns. Restrictions are specified by either (a) fixing parameters to a known value a priori (e.g., requiring that a variable is unrelated to a factor by fixing its regression in Λ to 0) or (b) constraining a set of two or more parameters to be equal. The equality constraints may be applied to any subset of parameters within or between groups, which provides the basis for specifying a model requiring invariant factor structures between multiple groups or across longitudinal occasions (as needed, for example, to test the hypothesis of measurement equivalence). Overidentified models (which have more restrictions than are necessary to identify the model parameters) place restrictions on the hypothesized form of Σ, which may be used to test the goodness of fit of the model to the data using the likelihood test statistic. Differences in chi-square between *nested* models (models that have the same specification, with additional restrictions in one model) may be used to test the null hypothesis that the restrictions (e.g., constrained equal factor loadings) are true in the population.

In multiple group, longitudinal factor analysis, it is necessary to estimate factor models using covariance metric and sample covariance matrices

Table 1
Reparameterized Sequential Sample for Multiple Group Analysis

Sample	Cohort (mean birth year)	Mean age			n
		0_1	0_2	0_3	
Group 1		30,	37,	44	109
1	1931	25,	32,	39	21
1	1924	32,	39,	46	26
2	1938	25,	32,	39	22
2	1931	32,	39,	46	40
Group 2		42,	49,	56	160
1	1917	39,	46,	53	27
1	1910	46,	53,	60	32
2	1924	39,	46,	53	51
2	1917	46,	53,	60	50
Group 3		58,	65,	72	143
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

Note. 0_1 = first occasion of measurement; 0_2 = second occasion of measurement; 0_3 = third occasion of measurement.

rather than to analyze separately standardized correlation matrices. Standardization could obscure invariant factor structures because of group differences in observed variances (Jöreskog, 1971), and would not allow evaluation of longitudinal changes in factor variances. To estimate raw score factor pattern weights and factor variances, one must identify the metric of the factors by fixing a single regression in each column of Λ to a constant (conveniently, 1.0), and then interpret results while considering the metric of latent and observed variables. The analyses reported here do so. Nevertheless, as standardized factor loadings (etc.) are easier to interpret, we provide parameter estimates that have been rescaled to a quasi-standardized metric, using a SAS PROC MATRIX program for scaling longitudinal factor analyses.[2] This rescaling preserves longitudinal constraints on parameter estimates but returns scaled values for factor loadings that are similar to standardized factor loadings. We also report maximum likelihood estimates and standard errors for certain models so that the reader may evaluate (a) a null hypothesis that each parameter is equal to zero, or (b) that group differences in unconstrained parameters are statistically reliable. In general, parameters that exceed their standard errors by a ratio of 2:1 are reliably different from zero at a 5% (per comparison) alpha level.

Results

The longitudinal models we estimate are designed to test the properties of the second-order general intelligence factor (*g*) from the PMA identified by Thurstone and Thurstone (1941). A first step was to determine that the *g* factor was an adequate representation of the covariance structure of the five PMA subtests. Bechtoldt (1974) and Corballis and Traub (1970) worked with a two-factor representation of the PMA subtests, although Bechtoldt's work included an additional memory variable that was not included in the 1948 PMA, and Corballis and Traub's two-factor model appeared to produce a very weak second factor. Nevertheless, we considered it necessary to evaluate the sufficiency of the *g* factor model before proceeding to longitudinal analysis. To do so, we used an exploratory factor analysis of all first-occasion cross-sectional data from the SLS (N = 2,202) to estimate an unrestricted maximum likelihood factor solution. The results for the one-factor model clearly indicated that the *g* factor sufficiently accounted for the covariance structure, $\chi^2(5, N = 2{,}202) = 6.18, p < .25$; Tucker-Lewis reliability = .997.

Longitudinal Model: Sample 1

Prior to analyzing the multiple age groups, we first analyzed the longitudinal factor model for the entire Sample 1. This analysis permitted us to evaluate the structural model prior to engaging in the more complex multiple group models reported later in the article. The basic occasion-specific model is depicted in Figure 1. The *g* factor was specified at each longitudinal occasion. The metric of *g* was defined by fixing the loading of Reasoning on *g* to 1.0. The remaining four factor loadings at each occasion were freely estimated, but were constrained to be equal across longitudinal occasions. By design, the loadings of all of the other variables (e.g., Verbal Meaning at Time 3 on *g* at Time 1) were fixed at 0. The factor covariance matrix was freely estimated, and the residual covariance matrix was specified as a diagonal matrix of unique variances.

We hypothesized in advance that this model would not fit the data because of the diagonal specification for the residual covariance matrix. It is well-known that longitudinal factor models of the type we are working with are likely to require what has been termed *autocorrelated* residuals (Sörbom, 1975; Wiley & Wiley, 1970). That is, given that it is likely that the occasion-specific factors will not account for all the reliable variance in the observed variables, then it is plausible to expect that the residuals (specific components) for an observed variable will correlate over time. In other words, we expected a residual covariance between the residual for Verbal Meaning at Time 1 and the Verbal Meaning residual at Time 2, a residual covariance between the Time 1 Space residual and the Time 2 Space residual, and so on. This residual pattern was especially likely, given that we are estimating a second-order *g* factor, as in this case the residual will include variance in the primary ability not accounted for by *g*. In fact, one would expect from the literature on abilities that the communalities for variables like Space and Number determined by *g* would be relatively small.

The initial model, denoted 0_1, specifying a diagonal matrix of unique variances provided an exceptionally poor fit to the data (see Table 2). The poor fit was underscored by the fact that the estimated factor covariances were greater than the corresponding factor variances (which implies the logical absurdity of correlations greater than 1). We therefore estimated Model 0_2, specifying autocorrelated residuals in the residual covariance matrix. The improvement in fit was substantial, change in $\chi^2(15, N = 162) = 898.64, p < .001$. Indeed, the overall chi-square test statistic was no longer significant, and the normed fit index was .96, indicating that nearly all the covariance in the sample data matrix was accounted for by the model.

At this point, our interest shifted to testing hypotheses regarding cross-occasion invariance in the parameter matrices. The principal hypothesis of interest with respect to measurement equivalence involved the invariance of the raw-score factor pattern weights (factor loadings) in Λ. Model 0_3 relaxed the constraint that the factor pattern weights be equal across occasions. The difference in fit was nonsignificant, indicating that the hypothesis of equal weights could not be rejected.

Given invariant factor pattern weights, it was meaningful to ask whether the factor variances were stationary over time, indicating consistency in the magnitude of individual differences on *g*. Model 0_4 tested this hypothesis by constraining the diagonal elements of the factor covariance matrix to be equal across longitudinal occasions. This hypothesis was rejected (see Table 2). Thus we concluded that there were changes in the magnitude of individual differences over occasions. We were also able to reject the null hypothesis that the factor covariances were equal (see Model 0_5 of Table 2).

Next, our attention turned to the parameters in the residual covariance matrix. Our first hypothesis was that the residual covariances could be constrained equal over occasions. This hypothesis, if tenable, would suggest a high degree of stability of individual differences in the ability-specific residual components. As can be seen in Table 2, Model 0_6, imposing the equality constraints on the residual covariances, did not fit worse than the Model 0_3, indicating that the hypothesis of equal covariances

[2] Briefly, the scaling is accomplished by pooling estimated latent variances and estimated observed variances to obtain scaling matrices. Pooling is done over multiple groups, as in Jöreskog (1971), and also over longitudinal occasions. A set of scaling equations and a listing of the scaling program is available from the first author on request.

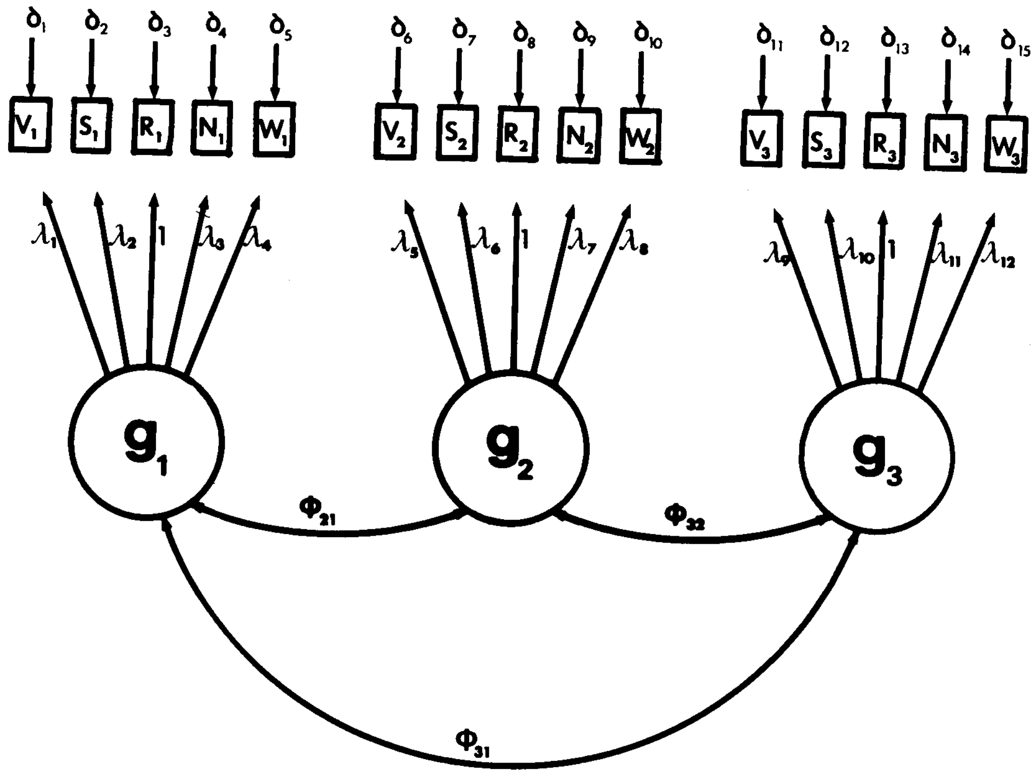

Figure 1. Initial longitudinal factor model specifying general intelligence factor (g) at each of three longitudinal occasions. (Subsequent models include covariances among corresponding residuals [e.g., δ_1, δ_6, δ_{11}] over time.)

could not be rejected. Finally, we tested the hypothesis of longitudinal invariance in the residual variances. This hypothesis stipulates that longitudinal changes in the variances of the observed variables could be attributed to changes in g factor variance alone. This model, labeled 0_7 in Table 2, was rejected as an equivalent representation to Model 0_6. We concluded that there were occasion-specific differences in the unique variances as well as in the factor variances.

The factor loadings their associated standard errors of the accepted model (0_6) are given in Table 3. All factor loadings are significant, but the rescaled factor loadings for Verbal Meaning and Reasoning are clearly larger than the rest. This pattern is consistent with the factor analytic literature on second-order ability factors (e.g., Horn, 1978), and parallels the findings of Thurstone and Thurstone (1941).

This pattern is also reflected in the standardized residual variances, where the smallest residuals (largest communalities) are associated with Verbal Meaning and Reasoning. Note also the

Table 2
Goodness-of-Fit Statistics for Alternative Longitudinal Models

Model	χ^2	*df*	*p*	ρ^a	Comparison	$\Delta\chi^2$	Δdf	*p*	$\Delta\rho$
$0_1(\Lambda_t =,^b \text{diag } \Theta^c)$	985.84	95	.000	.574	—	—	—	—	—
$0_2(\Lambda_t =, \text{cov } \Theta^d)$	87.20	80	.27	.962	0_1–0_2	898.64	15	< .001	.388
$0_3(\Lambda_t \neq)$	82.98	72	.17	.964	0_1–0_3	4.22	8	*ns*	.002
$0_4(\Lambda_t =, \text{diag } \Phi_t =^e)$	112.90	82	.013	.951	0_4–0_2	25.70	2	< .001	.011
$0_5(\Lambda_t = \Phi_t =^f)$	121.78	84	.005	.947	0_5–0_4	8.88	2	< .05	.004
$0_6(\Lambda_t =, \text{cov } \Theta =^g)$	97.16	90	.28	.958	0_6–0_2	9.96	10	*ns*	.004
$0_7(\Lambda_t = \Theta_t =^h)$	129.21	100	.026	.944	0_7–0_6	32.05	10	< .05	.014

[a] Bentler–Bonett normed fit index.
[b] Indicates nonzero factor pattern weights in Λ constrained to be equal over time (t).
[c] Indicates the residuals in Θ specified as uncorrelated (see text).
[d] Indicates autocorrelated residuals in Θ. This specification was continued in Models 0_3–0_7, as well.
[e] Indicates factor variances in Φ constrained to be equal over time.
[f] Indicates factor covariances constrained equal, and factor variances constrained equal over time.
[g] Indicates covariances among residuals constrained equal over time.
[h] Indicates residual variances constrained equal over time, and residual covariances constrained equal over time.

Table 3
Factor Loadings and Residual Variances for the Longitudinal Factor Model (0_6)

Test	Factor loadings		Residual variances[a]		
	LISREL estimates[b]	Rescaled loadings	Time 1	Time 2	Time 3
Verbal Meaning	1.540 (0.100)	.838	.318	.348	.240
Space	0.994 (0.109)	.556	.751	.666	.652
Reasoning	1.00[c] (—)	.878	.269	.274	.162
Number	0.928 (0.108)	.518	.760	.763	.674
Word Fluency	1.108 (0.133)	.520	.774	.735	.682

[a] Calculated as the proportion of residual variance (estimated) to total variance (estimated); 1 − (residual variance) = the communality.
[b] Standard errors in parentheses.
[c] Fixed parameter.

longitudinal decreases in residual variances for all variables, suggesting that the communalities of the primary ability variables determined by g increase over time. The high degree of stability in individual differences is reflected in the high factor covariances, which are provided in Table 4. Standardized, these covariances reflect correlations of greater than .9 between g at each longitudinal occasion. Clearly, there is not much change in the relative ordering of individuals on general intelligence over the 14-year period.

The results of this model were successfully cross validated in Sample 2. Rather than report these results, we move immediately to discussion of the multiple group analysis.

Multiple Group Analysis

The analyses in Samples 1 and 2 suggest almost perfect stability of individual differences in intelligence, both at the g factor and test-specific component levels. These analyses combined individuals spanning the adult life span, however, and it was possible that the wide age range served to maximize the apparent stability of individual differences. In particular, it was possible that differential change in the late-middle-age/old-age ranges was obscured by the high degree of stability across most of the adult life span. The multiple group analyses were designed to examine the stability of individual differences in more homogeneous age ranges. They also afforded us the opportunity of looking at age group differences in the factor analysis parameters. One might expect that there would be a greater opportunity for age group differences in factor loadings—given the age ranges spanned by our groups—than for longitudinal age changes.

We began by testing the equality of the observed covariance matrices across the three age groups. Box's test suggested non-homogeneous covariance matrices, $M = 402.77$, $F(240, \infty) = 1.59$, $p < .0001$. This result made it likely that there indeed were group differences in some of the factor analytic parameters.

Table 4
Factor Covariance Matrix (and Correlations) for the Longitudinal Factor Model (0_6)

Factor	g_1	g_2	g_3
g_1	28.624 (4.137)	0.945	0.917
g_2	27.723 (3.983)	30.062 (4.338)	0.972
g_3	31.776 (4.531)	34.528 (4.787)	41.938 (5.728)

Note. g_1 is the general factor at Time 1, g_2 is the general factor at Time 2, g_3 is the general factor at Time 3. Standard errors in parentheses. Values above the diagonal are standardized factor correlations.

The longitudinal factor model investigated in Sample 1 was used in the multiple group analyses. However, rather than presume the equivalence of residual covariances (as in Model 0_6 above) we chose to begin with these parameters unconstrained. Our rationale was that group differences in the residual covariance structure might have been obscured in the single sample analysis. Rather than presume the constraints, we chose to evaluate them anew in the multiple group model.

Our basic model, then, posited the specification of Model 0_3 of the Sample 1 analyses: an occasion-specific g factor (with no longitudinal constraints on the factor loadings), a freely estimated factor covariance matrix, and a residual covariance matrix with free unique variances and autocorrelated residual covariances. This model was specified in each of the three age groups, with no additional constraints on the parameters across the groups. The model was therefore equivalent to running the longitudinal factor model separately in the three groups.

As can be seen from the first entry in Table 5, this model denoted M_1, provided a relatively good fit to the data, allowing us to conclude that it was a reasonable representation of the covariance matrices in each group. We therefore proceeded to test for invariance in the g factor loadings. Separate tests of the equality of the factor loadings across age groups (Model M_2) and longitudinally across occasions (Model M_3) did not fit worse than the model with no constraints on the factor loadings (see Table 5). For both tests, the combined change in chi-square was actually just less than the change in degrees of freedom, $\chi^2(32, N = 412) = 29.82$, *ns*. We therefore concluded that the g factor loadings demonstrated complete age equivalence—being invariant both longitudinally and between age groups.

Our next set of models examined invariance in the factor covariance matrix. Model M_4, requiring age group equivalence in the factor covariances matrix (both variances and covariances), significantly degraded the fit to the data, requiring rejection of the null hypothesis of age group equivalence. We next tested a less restrictive model, positing group equivalence in factor variances but not in covariances. This model (M_5) was also rejected. Finally, Model M_6, placing no group constraints on the variances but positing longitudinal equality of variances within each group, was also rejected by the data (see Table 5). We should note that none of these models greatly degraded the fit, as judged by the normed fit index change of .01 or less (see Bentler & Bonett,

Table 5
Goodness-of-Fit Statistics for Models With Multiple Groups

Model	χ^2	*df*	*p*	ρ[a]	Comparison	$\Delta\chi^2$	Δdf	*p*	$\Delta\rho$
M_1 (all free)[b]	257.85	216	.027	.951	—	—	—	—	—
$M_2(\Lambda_g =$[c]$)$	284.24	240	.026	.946	M_2–M_1	26.39	24	*ns*	.005
$M_3(\Lambda_t =$[d]$)$	287.68	248	.042	.945	M_3–M_2	3.44	8	*ns*	.001
$M_4(\phi_g =$[e]$)$	329.65	260	.002	.937	M_4–M_3	41.97	12	< .01	.008
$M_5(\mathrm{var}\phi_g =$[f]$)$	310.68	254	.004	.941	M_5–M_3	23.00	6	< .01	.004
$M_6(\mathrm{var}\phi_t =$[g]$)$	301.28	254	.022	.943	M_6–M_3	14.00	6	< .05	.002
$M_7(\Theta_g =)$[h]	458.85	308	.000	.913	M_7–M_3	171.17	60	< .001	.032
$M_8(\mathrm{cov}\Theta_t =)$[i]	331.77	278	.015	.937	M_8–M_3	44.09	30	< .05	.008

[a] Bentler–Bonett normed fit index.
[b] Indicates no between-groups equality constraints among parameters.
[c] Indicates factor loadings constrained equal between groups.
[d] Indicates factor loadings constrained equal between groups (as in M_2) *and* constrained equal over time (this specification maintained in Models M_4–M_8).
[e] Indicates factor covariance matrices constrained equal between groups.
[f] Indicates factor variances constrained equal over groups.
[g] Indicates factor variances constrained equal over time in each of the groups.
[h] Indicates entire residual covariance matrix constrained equal over groups.
[i] Indicates residual covariances for test-specific components constrained equal over time.

1980). Nevertheless, the loss of fit, judged from the likelihood ratio chi-square test, was significant. These results indicated that the factor covariance matrices should neither be taken to be stationary over time nor equivalent across age groups.

Finally, we pursued the residual covariance structure to assess the stability of the residual variances and covariances across time. A preliminary model, M_7, specified group invariance in all three parameter matrices (Λ, Φ, and Θ). Compared to model M_3, this model tests the age group equivalence of the residual covariance matrix. The hypothesis was convincingly rejected. Our next step was to evaluate the plausibility of a model constraining the residual covariances to be equal between different measurement occasions (as was the case for Model 0_6 in the single sample analysis). Model M_8 placed these constraints on the residuals. The loss of fit was marginally significant at the 95% confidence level. We concluded that the model specifying equal covariances had missed the mark, but not by much. Thus, unlike Model 0_6, we could not treat the residual covariances as invariant over longitudinal occasions in the multiple group analysis. Apparently, both the residual variances and covariances differed by group and over longitudinal occasions, although the loss of fit due to group constraints was clearly much greater than the loss due to fitting invariant residual covariances over longitudinal occasions in each of the groups separately.

An alternative method for approaching stability in the residual covariances is by specification of a model positing both occasion-specific and test-specific factors (e.g., Jöreskog & Sörbom, 1977). Figure 2 depicts the factor pattern matrix (Λ) associated with a combined occasion-specific and test-specific factor model for these data. A given variable loads both on the general factor and its own test-specific factor (i.e., a Verbal Meaning factor, a Space factor, and so on). This parameterization of the residual covariances is plausible if one argues for a special relation among the residuals over time—a first order autoregressive structure (see Jöreskog & Sörbom, 1977). Addition of test-specific factors places no additional restrictions on the residual covariances, given that there are only three occasions of measurement (with more occasions, specification that the residual covariances form a single common factor may not fit the residual covariance structure). The advantage of the test-specific factor representation is that it enables one to separately estimate components of variance associated with *g*, stable variance in the primary ability, and a residual consisting of unstable variance plus measurement error (see Hertzog, in press).

We reestimated model M_3 (invariant factor loadings only) with test-specific factors. The parameter estimates and standard errors are provided in Tables 6 and 7. Given the fact that the hypothesis of invariant *g* factor loadings had been found plausible, we were entitled to assume measurement equivalence and to evaluate the remaining parameter estimates with respect to the issue of stability and change in intelligence. Several points of interest regarding the stability of individual differences emerged. First, the factor covariances were again extremely high, indicating a great degree of stability in individual differences in *g* over the 14-year interval for all three age groups. Standardized, these factor correlations are approximately .9 (or greater) for all groups (see Table 7).

Table 8 summarizes the stability of individual differences by reporting the correlations, r^2, and the estimated autoregressive coefficients predicting *g* from the previous longitudinal occasion. As can be seen from Table 9, the r^2 is larger for g_2 to g_3 in all groups, accounting for 92% of the variance in g_3 in both the middle-aged and old groups. The predominance of stability is underscored by the regression coefficients reported in Table 9. As suggested by Kessler and Greenberg (1981), we have expressed the raw-score slope coefficients in terms of the stability and, as given in the last column of Table 9, the regression of the change scores on initial scores (e.g., the regression of g_3–g_2 on g_2). This latter coefficient, if negative, suggests regression to the mean; if positive, it suggests increasing differences between individuals that covary with initial differences. Table 6 shows that the raw-score slopes were very near 1.0 (suggesting high stability) and that the change slopes were near zero (suggesting little change variance predictable from initial scores). In both the middle-

FACTORS

VARIABLES	g_1	g_2	g_3	V	S	R	N	W
V_1	λ_1	0	0	λ_5	0	0	0	0
S_1	λ_2	0	0	0	λ_7	0	0	0
R_1	1	0	0	0	0	λ_9	0	0
N_1	λ_3	0	0	0	0	0	λ_{11}	0
W_1	λ_4	0	0	0	0	0	0	λ_{13}
V_2	0	λ_1	0	1	0	0	0	0
S_2	0	λ_2	0	0	1	0	0	0
R_2	0	1	0	0	0	1	0	0
N_2	0	λ_3	0	0	0	0	1	0
W_2	0	λ_4	0	0	0	0	0	1
V_3	0	0	λ_1	λ_6	0	0	0	0
S_3	0	0	λ_2	0	λ_8	0	0	0
R_3	0	0	1	0	0	λ_{10}	0	0
N_3	0	0	λ_3	0	0	0	λ_{12}	0
W_3	0	0	λ_4	0	0	0	0	λ_{14}

Figure 2. Factor pattern matrix for model including occasion-specific and test-specific factors. (0's and 1's are fixed parameters; λ's are estimated by the model.)

aged and old groups, the change slopes were slightly negative for g_1 and g_2, suggesting slight regression to the mean, and slightly positive from g_2 to g_3, suggesting some egression from the mean (the rich getting richer, the poor poorer, as it were). In the young group, the stabilities were lower, albeit still impressively large, and the regression to the mean was consistent across time intervals.

The patterns of stability and change identified in the regression coefficients were mirrored in the factor variances, which exhibited different patterns of change across each of the groups. Factor variances decreased in the young group, but showed reliable increases from the second to the third occasion of measurement in both the middle-aged and old groups. This increase in *g* variance was consistent with the regression from the mean suggested from the regression coefficients. The decreases in variance and the regression to the mean pattern in the young group may reflect the mild ceiling effects on Verbal Meaning and Reasoning that we have observed in the youngest age groups in the SLS longitudinal samples.

Third, factor variances varied in magnitude between the age groups. The older group was generally more heterogeneous (had greater individual differences in *g*) than were the young and middle-aged groups. Taken together, these results suggested that although there was significant stability of individual differences in all age groups, the old group showed an interesting pattern of (a) greater variability in *g* at initial measurement and (b) increasing variability over time.[3]

An alternative way of looking at stability is the decomposition of variance in the model including both occasion-specific and test-specific factors. As can be seen in Table 9, the preponderance of *g* variance at the second and third occasions of measurement is stable variance predicted by individual differences at the prior measurement occasion. Given that we were studying the second-order *g* factor, it is relevant to ask about the stability of the residual components, reflecting the five primary ability factors from the PMA. Table 9 reports the decomposition of variance on each of the 15 observed variables for each group into proportions of (a) *g*-related variance, (b) stable test-specific variance, and (c) residual variance. The *g*-related variance components are actually the communalities of the observed variables with respect to the *g*

[3] One concern we had was that the patterns of factor variances might be due to the different age span for the oldest group (see Table 1). We therefore reanalyzed the data, using only the two oldest cohorts in Samples 1 and 2 to form a smaller old group. The redefinition of the old group did not eliminate the higher variances in *g* for the old, but did attenuate the longitudinal increases in variance. This analysis is discussed in more detail in the second article in this series (Hertzog & Schaie, 1986).

Table 6
Factor Loadings for Model With Occasion-Specific (g) and Test-Specific Factors

Variable	g[a]	g^*[b]	Test (Young)[c]	Test (Middle aged)[c]	Test (Old)[c]
V_1	1.659 (.098)	.767	1.032 (.129)	0.921 (.122)	0.650 (.193)
S_1	0.948 (.087)	.438	1.001 (.084)	0.908 (.107)	1.136 (.208)
R_1	1.000* —	.777	0.752 (.174)	1.120 (.151)	0.708 (.199)
N_1	1.463 (.106)	.588	1.005 (.086)	0.962 (.058)	0.935 (.084)
W_1	1.340 (.118)	.485	0.667 (.102)	1.049 (.102)	1.046 (.104)
V_2	1.659 (.098)	.767	1.000* —	1.000* —	1.000* —
S_2	0.948 (.087)	.438	1.000* —	1.000* —	1.000* —
R_2	1.000* —	.777	1.000* —	1.000* —	1.000* —
N_2	1.463 (.106)	.588	1.000* —	1.000* —	1.000* —
W_2	1.340 (.118)	.485	1.000* —	1.000* —	1.000* —
V_3	1.659 (.098)	.767	0.971 (.120)	0.820 (.117)	1.042 (.323)
S_3	0.948 (.087)	.438	0.965 (.089)	0.770 (.095)	1.130 (.211)
R_3	1.000* —	.777	0.920 (.208)	1.006 (.133)	0.740 (.196)
N_3	1.463 (.106)	.588	0.970 (.080)	0.868 (.053)	0.786 (.074)
W_3	1.340 (.118)	.485	0.988 (.126)	0.925 (.086)	0.928 (.092)

Note. Standard errors are in parentheses. Asterisks denote fixed parameters. Subscripts on variables indicate longitudinal occasion (1 = Time 1, 2 = Time 2, 3 = Time 3). V = Verbal Meaning; S = Space; R = Reasoning; N = Number; W = Word Fluency.
[a] Factor loadings for occasion-specific general factor (g). Estimates were constrained equal across the 3 longitudinal occasions.
[b] Rescaled general factor loadings.
[c] Test-specific factor loadings for each age group.

factor. The variance associated with the test-specific factor represents stable variance across occasions specific to the primary ability. The residual variance represents a combination of measurement error variance and unstable specific variance (the two components cannot be disentangled in this analysis). There are several points of interest in Table 9. First, the communalities of the g factor increased substantially in the old group relative to the young and middle-aged groups (and showed a tendency to increase over time longitudinally as well). Thus g determines more of the variance of the observed measures in the old than in the young. Second, those variables with the lowest communalities for g (Space, Number, Word Fluency) show very high levels of stability in the primary ability (test-specific) domain. For example, although only about 14% of the young group's variance of Space at Time 1 is determined by g, 72% of Space's Time 1 variance is determined by the Space test-specific factor in the young group. This indicates substantial stability in both the g and test-specific domains. Proportions of stable test-specific variance to total g-adjusted variance are given in the right-hand column of Table 9. Considering that these proportions are contaminated by measurement error, the proportion of stable variance in the primary ability measures independent of g is indeed impressive. Finally, the unique variances show some evidence of change in the primary abilities, but in many cases the proportions of unique variance are close to what would be expected to be the magnitude of error variance, given the reliabilities of the measures reported by Thurstone and Thurstone (1949).

Table 7
Factor Covariance Matrices for Occasion-Specific Factors in Each Age Group

Factor	g_1	g_2	g_3
	Young		
g_1	15.048 (2.868)	0.887	0.930
g_2	11.896 (2.409)	11.959 (2.421)	0.933
g_3	11.951 (2.365)	10.690 (2.179)	10.970 (2.257)
	Middle aged		
g_1	16.797 (2.691)	0.927	0.960
g_2	16.204 (2.549)	16.761 (2.652)	0.959
g_3	16.786 (2.607)	16.760 (2.591)	18.204 (2.798)
	Old		
g_1	23.546 (3.595)	0.944	0.885
g_2	22.405 (3.427)	23.941 (3.713)	0.959
g_3	23.442 (3.598)	25.589 (3.814)	29.769 (4.335)

Note. Standard errors are in parentheses. Values above the diagonal are factor correlations, standardized independently in each age group.

Table 8
Correlations and Regression Coefficients Indicating Stability of Individual Differences in g

Group	r[a]	r^2	$1-r^2$	b[b]	$b_{\Delta x}$[c]
Young					
g_1, g_2	.887	.787	.213	0.791	−0.209
g_2, g_3	.933	.870	.130	0.894	−0.106
Middle aged					
g_1, g_2	.927	.859	.141	0.965	−0.035
g_2, g_3	.959	.920	.080	1.000	0.000
Old					
g_1, g_2	.944	.891	.109	0.952	−0.048
g_2, g_3	.959	.920	.080	1.069	0.069

Note. Stabilities are shown for 7-year intervals between adjacent longitudinal occasions.
[a] Simple correlation of scores for adjacent occasions.
[b] Simple regression of later occasion on earlier occasion (unstandardized).
[c] Regression of change score on earlier occasion (unstandardized).

Table 9
Estimated Variance Components From Final Multiple Groups Model

Variable	$\hat{\sigma}^2$	g[a]	Test-specific[b]	Unique[c]	Stable (test)[d]
			Young		
V_1	76.286	.543	.333	.124	.729
S_1	98.688	.137	.724	.139	.839
R_1	27.518	.547	.186	.268	.410
N_1	114.107	.282	.591	.127	.823
W_1	136.097	.199	.332	.470	.414
V_2	62.008	.531	.385	.084	.821
S_2	103.512	.104	.689	.208	.768
R_2	27.832	.430	.325	.246	.569
N_2	115.828	.221	.576	.203	.739
W_2	148.667	.144	.683	.173	.798
V_3	63.587	.475	.354	.171	.674
S_3	104.872	.094	.634	.274	.698
R_3	24.025	.457	.318	.225	.586
N_3	96.291	.244	.652	.104	.862
W_3	159.879	.123	.620	.257	.713
			Middle aged		
V_1	77.203	.599	.273	.128	.681
S_1	97.719	.153	.468	.369	.559
R_1	32.471	.517	.299	.184	.448
N_1	120.387	.299	.589	.112	.840
W_1	154.861	.195	.502	.304	.623
V_2	81.848	.564	.304	.133	.696
S_2	82.420	.183	.680	.137	.832
R_2	29.076	.576	.266	.157	.629
N_2	127.997	.280	.599	.121	.832
W_2	125.227	.240	.564	.196	.742
V_3	81.260	.617	.206	.178	.536
S_3	85.797	.191	.387	.422	.478
R_3	30.568	.596	.256	.148	.634
N_3	109.363	.356	.528	.116	.820
W_3	119.517	.273	.505	.221	.696
			Old		
V_1	102.167	.634	.087	.278	.238
S_1	83.784	.253	.348	.400	.465
R_1	34.374	.685	.105	.210	.333
N_1	119.696	.421	.441	.138	.762
W_1	163.680	.258	.516	.226	.695
V_2	115.064	.573	.184	.243	.431
S_2	77.426	.278	.292	.431	.404
R_2	36.005	.665	.201	.134	.600
N_2	129.347	.396	.466	.138	.772
W_2	152.787	.281	.505	.214	.702
V_3	126.724	.647	.182	.172	.514
S_3	74.825	.356	.385	.258	.599
R_3	38.027	.783	.104	.113	.479
N_3	119.211	.534	.313	.153	.672
W_3	151.523	.353	.438	.208	.678

Note. $\hat{\sigma}^2$ = estimated variance of observed variable. V = verbal meaning; S = space; R = reasoning; N = number; W = word fluency. Subscripts on variables indicate longitudinal occasion (1 = Time 1, 2 = Time 2, 3 = Time 3).
[a] Proportion of variance due to g.
[b] Proportion of variance due to test-specific factor.
[c] Proportion of variance unique to the observed variable. The sum of the three proportions (g-related, test-specific, unique) is 1.0.
[d] Proportion of variance *not* determined by g that is determined by the test-specific factor.

Discussion

The results of the present study present a relatively coherent picture—one of measurement equivalence and stability in psychometric intelligence, as measured by the Thurstones's 1948 Primary Mental Abilities test, in adulthood. We found that it was highly plausible to model the factor loadings of a general intelligence factor as being invariant, both longitudinally and across multiple age groups. We also found a high degree of stability of individual differences across the adult life span.

The finding of invariance in the g factor loadings is important relative to the suggestion in the literature that the fundamental measurement properties of the psychometric tests change over the life span (e.g., Baltes & Nesselroade, 1970; Demming & Pressey, 1957; Schaie, 1977). As shown by Meredith (1964), under selection of subpopulations from a population for which an isomorphic common factor model holds, the multiple subpopulations will have an invariant unstandardized factor pattern matrix. Meredith's work implies that one must reject the hypothesis of metric invariance before one is justified in concluding that the groups have qualitatively different factor structures. One cannot argue for qualitative group differences in measurement properties if the hypothesis of metric invariance cannot be rejected. In contrast, we found the hypothesis of metric invariance to be strongly supported by our data. Our results therefore suggest that, whatever the faults inherent in the constructs of psychometric intelligence, measures of psychometric intelligence seem to be measuring basically isomorphic constructs with similar measurement properties at different age levels.

One could still, of course, argue that the constructs measured by psychometric intelligence are of limited utility in predicting intelligent behaviors in adults (e.g., Sternberg, 1985). Nevertheless, our findings do not support the notion that psychometric testing of abilities in older populations is invalid because one is measuring qualitatively different constructs with unstable measures. Our conclusion must be qualified by the fact that our assessment of factorial invariance is specific to the second-order g factor. We cannot assess the invariance of the primary ability factor loadings from our data. We therefore cannot rule out the possibility of nonequivalent measurement properties at the primary ability level, although, given the stability indicated by the test-specific factors, the likelihood of measurement equivalence in the primary ability factors seems quite high. Data we recently collected on an expanded ability battery as part of the 1984 SLS assessment should help us address the measurement equivalence issue at the primary ability factor level.

The finding of factorial invariance is relevant to the factor analytic literature suggesting de-differentiation of ability factors in old age (Reinert, 1970). The de-differentiation argument states that ability factors coalesce, or collapse, toward a general intelligence factor in older groups. The early literature on this phenomenon was plagued by methodological inadequacies (Cunningham, 1978; Reinert, 1970; Schaie & Hertzog, 1985). Recent comparative factor analysis work by Cunningham (1980, 1981), using confirmatory factor analysis methods, suggests that there is little evidence for gross collapse of the factor space—the same number of factors are needed to model ability variables in old groups, and the loading patterns are highly similar. Our results are consistent with Cunningham's findings in suggesting invariance in the raw-score regressions of variables on ability factors, both across age groups and longitudinally within age groups (see also Cunningham & Birren, 1980).

Cunningham (1980, 1981) reported evidence for a mild form of de-differentiation—that is, increased factor correlations in the older groups. Our finding of increased communalities for *g* in the old group is also consistent with this mild form of de-differentiation. To clarify the relation, we report in Table 10 correlations among the primary abilities obtained by a confirmatory factor analysis specifying test-specific factors. As can be seen in Table 10, there is a pronounced tendency for factor correlations to be higher in the old group. Crude indexes of this tendency are the average correlations of .36 for the young group, .39 for the middle-aged group, and .54 for the old group. Nevertheless, it must be emphasized that the primary thrust of the de-differentiation argument—qualitative change in the nature of ability factors—is neither supported by Cunningham's findings nor by our own.

The age-related measurement equivalence in the PMA allows us to make unambiguous interpretation of the stability of individual differences in *g* over time. Clearly, individual differences in general intelligence are highly stable across 14-year longitudinal epochs for three age groups (spanning most of the adult age range). The stability coefficients indicated that approximately 90% of the *g* variance in the middle-aged and old groups was consistent between adjacent 7-year testing intervals. There is, then, little indication in these data of any substantial degree of variability in developmental trajectories in *g*. Moreover, the stability of individual differences in the PMA ability-specific components in our longitudinal model suggest a high degree of stability in individual differences on the primary abilities as well.

Although these results clearly limit the degree to which one could argue for a substantial degree of interindividual differences in intraindividual change in psychometric intelligence in adulthood, it would be overstating the case to argue that these data demonstrate a *lack* of variability in change functions across the adult life span. For one thing, it is well-known that the longitudinal samples of the SLS are influenced by a substantial degree of experimental mortality (Schaie, Labouvie, & Barrett, 1973), causing the participants in the 14-year studies to be relatively select with respect to ability levels. It is highly likely, given the relatively long 7-year retest interval and the nature of the sampling procedures, that individuals in terminal decline or suffering differential loss of abilities due to severe illness will have dropped out of the longitudinal sample (Hertzog, Schaie, & Gribbin, 1978). The high degree of stability we observed in this study may be specific to more select, healthy subpopulations of adults and may not generalize to the population at large. Moreover, our sample size was sufficiently small that we were forced to pool over relatively large age ranges to form our age groups. Such a procedure maximizes individual differences at the initial measurement occasion and may have obscured some degree of heterogeneity in developmental trends. We note, however, that the estimates of stability did not differ greatly between the Sample 1 analysis and the age-partitioned multiple group analysis that reduced individual differences produced by wide age spans.

Of course, as McCall (1981) pointed out, even stabilities of .9 allow for a greater degree of crossover of individual curves than might be expected by social scientists. At the individual level, it is still possible that a given individual will buck the tide, and exhibit less change in *g* than his or her same-age peers. There may also be more variability in the primary abilities than in the higher order intelligence factor. One can see in Table 7 that the test-specific stabilities were in some cases smaller than the stabilities for *g* in the same age interval. In the old group, for example, the stability of the Space test-specific factor seems to be smaller than the stability observed for Space in the young and middle aged, even though the stability of individual differences in *g* is, if anything, greater in the old group. This result may indicate slightly more variability in the patterns for the Spatial Orientation ability tapped by the Space test (see McGee, 1979). These data are not optimally suited for assessing primary ability-specific change, however, because unreliability due to measurement error cannot be separated from instability in the ability in the analysis we have reported. In any case, we must be careful to emphasize that there is considerably more consistency than inconsistency in age changes in all age groups, and for all PMA subtests. Finally, we cannot rule out the possibility that individual differences in change (or for that matter, changes in factor loadings), occur in older ages (beyond 80) not represented in this study.

The invariance in the PMA *g* factor loadings and the stability of individual differences in intelligence contrasts sharply with patterns of mean age changes found in the SLS (e.g., Schaie, 1983; Schaie & Hertzog, 1983). Schaie has consistently found variation in mean patterns according to age, cohort, and time of measurement. Moreover, these mean changes have been found to vary in magnitude for different abilities. The difference in findings underscores the critical distinction between stability in means (i.e., on average, no age changes) and stability of individual differences. In normally distributed variables, stability of the means and stability of individual differences (as measured by

Table 10
Primary Ability Factor Correlations for the Three Age Groups

Variable	Verbal Meaning	Space	Reasoning	Number	Word Fluency
Young (M age = 37)					
Verbal Meaning	1				
Space	.115	1			
Reasoning	.559	.455	1		
Number	.390	.239	.489	1	
Word Fluency	.531	.034	.425	.334	1
Middle aged (M age = 49)					
Verbal Meaning	1				
Space	.296	1			
Reasoning	.711	.479	1		
Number	.419	.248	.441	1	
Word Fluency	.508	.039	.439	.308	1
Old (M age = 65)					
Verbal Meaning	1				
Space	.593	1			
Reasoning	.838	.650	1		
Number	.666	.528	.627	1	
Word Fluency	.557	.290	.202	.450	1

covariances) are statistically (and conceptually) independent. As one can see in the next article in this series (Hertzog & Schaie, 1986), we can observe stability of individual differences either when there are no mean age changes or when there are substantial mean changes over a given portion of the life span.

References

Baltes, P. B., Dittmann-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 behaviors* (Vol. 6, pp. 34–76). New York: Academic Press.

Baltes, P. B., & Nesselroade, J. R. (1970). Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology, 1,* 162–168.

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 adulthood and old age: Where Horn and Donaldson fail. *American Psychologist, 31,* 720–725.

Bechtoldt, H. P. (1974). A confirmatory analysis of the factor stability hypothesis. *Psychometrika, 39,* 319–326.

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

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.

Corballis, M. C. (1973). A factor model for analyzing change. *British Journal of Mathematical and Statistical Psychology, 26,* 90–97.

Corballis, M. C., & Traub, R. E. (1970). Longitudinal factor analysis. *Psychometrika, 35,* 79–93.

Cunningham, W. R. (1978). Principles for identifying structural differences: Some methodological issues related to comparative factor analysis. *Journal of Gerontology, 33,* 82–86.

Cunningham, W. R. (1980). Age comparative factor analysis of ability variables in adulthood and old age. *Intelligence, 4,* 133–149.

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

Cunningham, W. R., & Birren, J. E. (1980). Age changes in the factor structure of intellectual abilities in adulthood and old age. *Educational and Psychological Measurement, 40,* 271–290.

Demming, J. A., & Pressey, S. L. (1957). Tests "indigenous" to the adult and older years. *Journal of Consulting Psychology, 4,* 144–148.

Eckensberger, L. (1973). Methodological issues of cross-cultural research in developmental psychology. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 43–64). New York: Academic Press.

Hertzog, C. (1985). An individual differences perspective: Implications for cognitive research in gerontology. *Research on Aging, 7,* 7–45.

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. 8). Hillsdale, NJ: Erlbaum.

Hertzog, C., & Schaie, K. W. (1986). *Stability and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures.* Unpublished manuscript, Georgia Institute of Technology, Atlanta.

Hertzog, C., Schaie, K. W., & Gribbin, K. (1978). Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology, 33,* 872–883.

Horn, J. L. (1978). Human ability systems. In P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 1, pp. 211–256). New York: Academic Press.

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 & J. Kagan (Eds.), *Constancy and change in human development* (pp. 445–529). Cambridge, MA: Harvard University Press.

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical/statistical modeling (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980s: 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.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.

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. (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., (1980). *Simultaneous analysis of longitudinal data from several cohorts.* (Research Rep. 80-5). Uppsala, Sweden: University of Uppsala, Department of Statistics.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V users guide.* Chicago: National Educational Resources.

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.

Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis: Models of quantitative change.* New York: Academic Press.

McCall, R. B. (1981). Nature–nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development, 52,* 1–12.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86,* 889–918.

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

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 throughout the human life span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 467–484). New York: Academic Press.

Schaie, K. W. (1977). Quasi-experimental research 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. (1979). The primary mental abilities in adulthood: An exploration in the development of psychometric intelligence. In P. B. Baltes & O. G. Brim (Eds.), *Life-span development and behavior:* (Vol. 2, pp. 67–115). New York: Academic Press.

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. (1982). Longitudinal methods. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 91–115). Englewood Cliffs, NJ: Prentice-Hall.

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., Labouvie, G. V., & Barrett, J. J. (1973). Selective attrition effects in a 14-year study of adult intelligence. *Journal of Gerontology, 28,* 328–334.

Schaie, K. W., Rosenthal, F., & Perlman, R. M. (1953). Differential mental deterioration of factorially "pure" functions in later maturity. *Journal of Gerontology, 8,* 191–196.

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

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

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.

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

Received May 3, 1985
Revision received October 21, 1985 ■