In R. J. Sternberg (Ed.), Advances in Human Intelligence (Vol. 3, pp. 79-118). Hillsdale, NJ: Elrbaum. (1986)

3 Toward a Comprehensive Model of Adult Intellectual Development: Contributions of the Seattle Longitudinal Study

K. Warner Schaie
Christopher Hertzog
The Pennsylvania State University

INTRODUCTION

The purpose of this chapter is to review briefly some of the crucial features of the *Seattle Longitudinal Study* (SLS) and to attempt to render an account of some of its contributions to the study of intellectual development over the adult life span. This effort, which has now spanned almost three decades, is still ongoing. Consistent with temporal trends in its field, the focus of the SLS has shifted from descriptive analysis of age changes in mean performance levels on psychometric tests to a broader attempt to measure and perhaps even modify individual differences in adult intellectual development.

We begin our presentation with a review of salient work on adult intellectual development and brief background on the contextual setting of the SLS. We then describe the research methods, termed *sequential strategies*, that emerged as part of the study. At the substantive level, attention focuses on age changes and cohort differences in mean levels of performance on the Primary Mental Abilities (PMA), as well as the studies of individual differences in intellectual change. We then discuss, without technical details, findings from our application of structural equation models to the analysis of intellectual development. Finally, we comment on recent work that includes efforts to reverse intellectual decline as well as applications of information-processing models to the problem of characterizing adult individual differences in PMA performance.

Some Comments on the History of Work on Adult Intellectual Development

Serious work in applied psychology virtually began with the investigation of intellectual competence. Early objectives may have ranged from the orderly removal of mentally retarded children from the public school classroom (Binet & Simon, 1905) to the demonstration of Darwinian characteristics by means of the study of individual differences (Galton, 1869). Indeed, Binet's definition of the construct of intelligence remains a classic guide for contemporary endeavors: "To judge well, to comprehend well, to reason well, these are the essentials for intelligence. A person may be a moron or an imbecile if he lacks judgment, but with judgment he could not be either" (Binet & Simon, 1905, p. 196).

Although early work on intelligence was concerned primarily with the acquisition of mental functions in early life (cf. Brooks & Weintraub, 1976), theorists soon began to attend to questions of intellectual development beyond childhood. Writers such as Hall (1922), Hollingsworth (1927), and Pressey (Pressey, Janney, & Kuhlen, 1939) raised questions concerned with the age of attaining peak performance level, the maintenance of transformation of intellectual structures, and the decremental changes that were thought to occur from midlife into old age.[1]

Soon enough, empirical work began to appear in response to those questions. In his original American standardization of the Binet Tests, Terman (1916) assumed that intellectual development reached a peak at 15 and remained level thereafter. But large-scale studies with the Army Alpha Intelligence Test (Yerkes, 1921) quickly suggested that average young adults crested at a mental age of 13! Other early studies questioned this inference. For instance, Jones and Conrad (1933), who examined most inhabitants of a New England community between the ages of 10 and 60, found that there were substantial age differences on some sub-tests of the Army Alpha but none on others. And Wechsler (1939), in his standardization studies for the Wechsler-Bellevue, found that growth of intelligence neither ceases in adolescence nor declines uniformly at older ages.

Interest in intelligence tests peaked soon after World War II with the near-universal adoption of the WAIS (Wechsler Adult Intelligence Scale) and its derivatives into clinical practice (cf. Matarazzo, 1972) and with the development of widely accepted classification tests (e.g., the DAT or GATB batteries; cf. Cronbach, 1970). Disenchantment set in, however, with widespread criticism of the misapplication of intelligence tests in education (e.g., Kamin, 1974). Clinicians, moreover, complained that profile analyses of tests had low validity

[1]This account is concerned primarily with the psychometric view of intelligence. The contributions of the Genevan school are recognized, but Piaget did not seriously consider adult intellectual development until the end of his work (1972). For extrapolations of the Genevan model to adulthood see Riegel (1975), Sinott (1981), or Schaie (1977/78).

and that information gained on intellectual status contributed little to the planning of therapeutic intervention.

These criticisms notwithstanding, the fact remains that omnibus measures of intelligence have been useful in predicting scholastic and vocational success, and specific ability measures have had utility in differential prediction in those instances where special abilities can be expected to be of importance (e.g., Horst, 1954). Although a reasonable argument could be envisaged for the proposition that motivational and other personality variables might have greater potency than intelligence in predicting adjustment in competence in midlife and old age, empirical evidence for such an argument is less than convincing.

When dealing with the elderly, in particular, it soon becomes apparent that the assessment of intellectual competence once again attains paramount importance. For example, questions such as who shall be retired for cause (in the absence of mandatory retirement at an early age), whether individuals retain sufficient competence for independent living, or whether individuals can continue to conserve and dispose of their property, all involve the assessment of intellectual function. Given that these issues are of societal importance, it becomes necessary to address the factual issues involved in the development of intelligence beyond adolescence. Intraindividual changes of a decremental nature need to be differentiated from interindividual differences that reflect the obsolescent functioning of older cohorts when compared to younger peers. Further, we must learn why some individuals show intellectual decrement in early adulthood and others maintain or increase levels of functioning on some ability variables well into old age.

The series of studies reviewed in this chapter attempts to address these issues by asking the following broad questions: (1) Does intelligence change uniformly or in different ability patterns? (2) At what age can age decrements be detected reliably, and what is the magnitude of such decrements? (3) What are the patterns of generational differences, and what is their magnitude? (4) What accounts for the individual differences in intraindividual change in intellectual abilities across the life span?

BACKGROUND OF THE SEATTLE LONGITUDINAL STUDY (SLS)

The origins of the SLS can be traced to work done as part of directed studies by the senior author while an undergraduate at Berkeley, under the supervision of Professor Read D. Tuddenham. Having been introduced to the basic writings of L. L. Thurstone (1938, 1941) and discovering that the Primary Mental Abilities (PMA) had not been explored beyond adolescence, the senior author realized that here was a possibly fruitful topic for systematic pursuit. A pilot study with the most advanced form of the PMA showed that factorial independence was retained in adulthood and, moreover, suggested ability-related differentials in Adult PMA performance (Schaie, Rosenthal, & Perlman, 1953). A dissertation at the University of Washington replicated this work on a wider portion of the adult life span and related intellectual functioning to rigidity-flexibility (Schaie, 1958). This dissertation became the base for the subsequent longitudinal and sequential studies.

The search for a suitable population frame for the SLS led to the Group Health Cooperative of Puget Sound, one of the first broadly-based health maintenance organizations. Established for reasons other than research on cognitive behavior, it provided a subject pool with known demographic characteristics. It also provided an economical means of retrieving former participants in the later longitudinal studies.

But perhaps the most significant stimulation leading to the conversion of a one-time cross-sectional inquiry into a series of longitudinal studies came from reading the early reports of longitudinal studies of individuals then reaching middle age, such as the papers by Bayley and Oden (1955), Jarvik, Kallmann, and Falek (1962), and Owens (1953, 1959). Taken together, these studies implied that maintenance of most intellectual abilities was the rule at least into middle age and, for some abilities, into later life. These findings contrasted with the earlier cross-sectional literature and the senior author's own data. To resolve this inconsistency, it seemed necessary to conduct a short-term longitudinal inquiry with a broad cross-sectional panel such as the one previously tested (Schaie, 1958).

The first longitudinal inquiry (conducted in 1963) mainly confirmed the inconsistency between cross-sectional and longitudinal data, even those obtained from a single panel. A theoretical model was therefore built that accounted for these discrepancies (Schaie, 1965). Subsequent replications of our findings occurred in 1970 and 1977, with a fifth wave during 1984/85. These replications have increased our understanding of cohort differences and have permitted us to follow individuals over increasingly longer periods of time (for some as long as 28 years). In order to discuss our substantive findings, we must first describe the basic design of the SLS.

SEQUENTIAL STRATEGIES FOR THE COLLECTION OF DEVELOPMENTAL DATA

Explication of a general development model (Schaie, 1965) that resulted from a close examination of our first longitudinal follow-up suggested limitations inherent in both single-cohort longitudinal and single time-of-measurement cross-sectional designs. Several alternative *sequential strategies* were therefore suggested (cf. Baltes, 1968; Schaie, 1965, 1973, 1977, 1984). The reader will gain a better understanding of the design of the SLS if we briefly summarize the nature of our sequential strategies and then describe the general sampling design of the SLS.

Disentangling Age Changes from Age Differences

The primary purpose of the sequential strategies was to attempt to disentangle the confounding of age and cohort in traditional cross-sectional and of age and socio-cultural change in traditional longitudinal data collections. The term "sequential" implies no more than that the required sampling frame include a sequence of samples taken across more than one measurement occasion. Sequential data collection strategies can best be understood by differentiating their closely related sampling and analysis designs (Schaie & Baltes, 1975). Sampling design refers to those cells of a cohort-by-age (time) matrix that are sampled in a given study. Analysis design refers to the manner in which these cells are then organized to analyze for the effects of Age, Cohort, and Time-of-measurement.[2] Figure 3.1 gives an example of a cohort-by-age matrix showing the various sequential designs. Figure 3.1 also illustrates the confounding of three parameters of interest. Age and Cohort appear as the rows and columns of the matrix, while Time is contained within the matrix cells.[3]

Sampling Designs. Two types of sampling designs may be distinguished. The first uses the same panel of individuals to fill in the matrix, as in traditional longitudinal studies. A second method, common in sociological research, selects independent random samples of individuals (each observed only once) from the same cohorts. The matrix in Fig. 3.1 could have been produced by either approach. These designs have also been called longitudinal and cross-sectional sequences (Baltes, 1968). A cross-sectional sequence typically involves the replication of a cross-sectional study with the age range of interest being assessed for at least two different time periods. Estimates for each age level are obtained across multiple cohorts, and each sample is measured only once. The longitudinal sequence, by contrast, represents the measurement of at least two cohorts over the same age range. Estimates from each cohort are obtained at two or more points in time. The critical difference between the two approaches is that only the longitudinal sequence can provide data that permit the evaluation of intraindividual change and interindividual differences in intraindividual change (Baltes & Nesselroade, 1979; Schaie & Hertzog, 1982).

[2] Our use of the term time-of-measurement is identical with the term period preferred in sociological research (e.g., Mason et al., 1973).

[3] The reader interested in the perennial debate on whether and how these effects should be unconfounded is referred to papers by Adam (1978), Glenn (1981), Horn and McArdle (1980), Mason et al. (1973), and Schaie and Hertzog (1982, 1984). The issues involved are quite complex, highly technical, and beyond the scope of this more substantively oriented chapter.

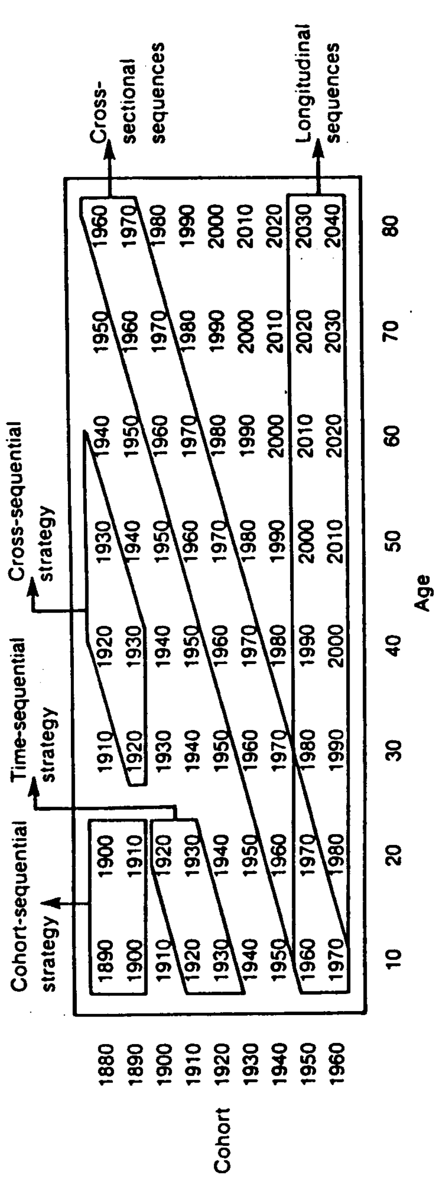

FIG. 3.1 Schematic showing cross-sectional and longitudinal sequences and the modes of analysis deducted from the general developmental model. Table entries represent time of measurement (from Schaie, 1983).

Analysis Designs. Data matrices such as that shown in Fig. 3.1 permit a variety of alternate analysis strategies (Schaie, 1965, 1973, 1977; Schaie & Hertzog, 1982, 1985). Each row of the matrix may be treated as a single-cohort longitudinal study, each diagonal as a cross-sectional study, and each column as a time-lag study.[4] Sequential sampling designs do not allow complete disentanglement of all three factors due to the obvious linear dependency between them. Nevertheless, there exist three distinct analysis designs that consider the separate effects of two of the components while assuming the constancy or irrelevance of the third on theoretical or empirical grounds.

The minimum designs of Fig. 3.1, show that the cohort-sequential will permit separation of age changes from cohort differences, given the assumption of trivial time-of-measurement effects. The time-sequential strategy permits the separation of age differences from time-of-measurement effects, given the assumption of trivial cohort effects. The cross-sequential strategy, finally, permits separation of cohort differences from time-of-measurement differences.

Analyzing Longitudinal Sequences. Both the cohort-sequential and the cross-sequential strategies can be applied to the analysis of longitudinal sequences. Many developmentalists find the cohort-sequential design of greatest interest because it explicitly differentiates intraindividual change within cohorts from interindividual differences between cohorts (cf. Baltes & Nesselroade, 1979; Schaie & Baltes, 1975). This design also allows an examination of the consistency of age functions across successive cohorts, thus providing an external validity check not available in the single-cohort longitudinal design.

As mentioned, a critical assumption for the validity of the cohort-sequential design (as well as for the traditional longitudinal method) is the absence of time-of-measurement effects in one's data. This assumption may be violated either by the presence of "true" secular trends or because of the occurrence of occasion-specific threats to internal validity such as differences in instrumentation or in experimenter behavior across measurement occasions. It is recognized by now that violation of the assumption justifying a specific sequential design will introduce inaccuracies into any estimates derived therefrom. The interpretational problem may be reduced by estimation of the relative likelihood of confounded T effects, given a strong theory about the nature and direction of estimated and confounded effects. To obtain such estimates, however, it may often be necessary to collect additional cross-sectional sequences.

Planned Violations of Assumptions. A pragmatically useful application of sequential designs can involve specification of an "invalid" design in order to obtain an indication of the existence of confounded effects. An example of such

[4] A time-lag study would involve the comparison of a behavior at a specific age for successive cohorts.

planned violation of assumptions would be the use of the cross-sequential (Cohort by Time) design under the assumption of no Age effects. In principle, the assumption is, on its face, unreasonable for the study of developmental phenomena. Such a design may nevertheless be useful when longitudinal data are available for a wide range of age/cohort groups but only for a limited number of measurement occasions. The cross-sequential design can be implemented after only two measurement occasions, whereas the cohort-sequential design requires a minimum of three. Further, the number of measurement occasions required for a cohort-sequential design that spans a wide age or cohort range would be prohibitive if we were to insist that no data analyses be performed until the entire cohort-sequential design desired for our research question was fully in hand. Given a strong theory about the nature of confounded Age effects, a cross-sequential design can provide useful information about the significance of the Age effects represented in the Cohort by Time design. Our early work on the sequential analysis of intelligence began with such misspecification in cross-sequential designs, in order to draw preliminary inferences regarding the relative importance of Cohort and Age effects prior to the availability of data that could have permitted direct simultaneous analyses (cf. Schaie & Strother, 1968). Nevertheless, our best estimates of age changes and generational differences have been obtained from more recent analyses of cohort-sequential designs, reported by Schaie and Hertzog (1983) and summarized later.

General Sampling Design of the SLS

Subject Selection. Our subjects are present and former members of the Group Health Cooperative of Puget Sound, a metropolitan area health maintenance organization. Our initial 1956 sampling frame consisted of approximately 18,000 potential subjects. These were stratified by age and sex, and 25 men and 25 women were randomly drawn from each birth year from 1880 to 1939. Of 2818 persons actually contacted, 910 agreed to participate, and data collection proceeded until 25 men and 25 women (total N = 500) had been tested in each five-year age interval from 21 to 70 (cf. Schaie, 1958).

In addition to the longitudinal follow-up during the 1963 cycle, 3000 names were again drawn randomly from the 1956 population frame after names of those tested in 1956 had been deleted. Of these, 996 persons over the age range from 22 to 77 years were tested. The third cycle, in 1970, followed a similar procedure. Survivors of the 1956 and 1963 panels were retested, and a new randomly selected panel of 705 members (aged 22 to 84 years) was initially tested. Our initial population frame having been virtually exhausted, we determined, by means of a collateral study, that it would be feasible to shift our population definition without significant impact upon the results from the study. For the fourth (1977) cycle, in addition to retesting survivors of the first three waves, we sampled approximately 3000 persons from what had now become a 210,000-member health plan. Of these, 705 persons (aged 22 to 84 years) were tested. During the 1984 fifth-wave data collection, we retested survivors of the first four waves and an additional 650-member panel.

Due to the standard seven-year interval between test occasions, all data are now organized in seven-year age and cohort groupings. For purposes of analysis, there are now ten different data sets, as follows:

1. A cross-sectional sequence consisting of four independent data sets: Aa (N = 500)—seven cohorts tested in 1956 (mean ages: 25 to 67; mean birth years: 1889 to 1938); Bb (N = 996)—eight cohorts tested in 1963 (mean ages: 25 to 74; mean birth years 1889 to 1938); Cc (N = 705)—nine cohorts tested in 1970 (mean ages: 25 to 81; mean birth years; 1889 to 1945); Dd (N = 609)—nine cohorts tested in 1977 (mean ages: 25 to 81; mean birth years 1896 to 1952).
2. Longitudinal sequences involving six data sets include three 7-year, two 14-year, and one 21-year follow-up: Ab (N = 303)—seven cohorts followed from 1956 to 1963; Bc (N = 420)—eight cohorts followed from 1963 to 1970; Cd (N = 340)—nine cohorts followed from 1970 to 1977; Ac (N = 162)—seven cohorts followed from 1956 to 1970; Bd (N = 337)—eight cohorts followed from 1963 to 1977; and Ad (N = 130)—seven cohorts followed from 1956 to 1977.

AGE CHANGES AND COHORT DIFFERENCES IN PMA MEANS

In this section we first describe the principal measurement variables used throughout our inquiries, then give a brief review of findings from the earlier ANOVA oriented analyses focusing on changes in ability level, and next summarize our most recent cohort-sequential analyses that feature theory-guided MANOVA approaches. We then comment on the implications for our findings of the current views on regression/structural approaches to age/cohort/time of measurement analysis.

The Measurement Variables

Whereas the Wechsler scales have been most widely used with adults for purposes of clinical diagnosis (cf. Matarazzo, 1972) and have also received substantial attention in developmental studies (cf. Botwinick, 1977; Schaie, 1980), they lack clear-cut factorial structure. Consequently, at the outset of our investigation, it seemed more reasonable to work with the factored tests provided by the work of the Thurstones (1938, 1941). The PMA test battery we used was derived from a series of factor-analytic studies of 56 mental tests by Thurstone (1938). The test form used in our studies was the 1948 PMA 11-17 version, which was, at the time our study began, the most difficult form of the test. It consists of five subtests designed to measure the abilities accounting for the greatest proportion of variance determined in the original factor analysis. All PMA subtests are known to have both substantial speed and power components (Anastasi & Drake, 1954). Indeed, the 1948 PMA Manual emphasizes the importance of the speed of thought as defining characteristics of intelligence (Thurstone & Thurstone, 1949). A description of the 1948 PMA subtests follows:

Verbal Meaning. This is the ability to understand ideas expressed in words. The test measures the range of a person's passive vocabulary in activities where information is obtained by reading or listening to words. The task requires synonym recognition in a multiple-choice format.

Space. This is the ability to think about objects in two or three dimensions. It may be described as the ability to imagine how an object or figure will look when it is rotated, to visualize objects in two or three dimensions, and to see the relations of an arrangement of objects in space. The test requires that the subject match geometric figures to a stimulus and mark those that are the same even though rotated, while not marking mirror images of the stimulus. This ability is often referred to technically as "spatial orientation."

Reasoning. This ability (in current ability factor taxonomies, often more specifically identified as "inductive reasoning") involves the solution of logical problems. According to Thurstone, persons with good reasoning ability can solve problems, foresee consequences, and analyze a situation on the basis of past experience. Reasoning is measured here by a letter series task, in which letter series are based on one or more rules and the problem is to discover the rule and mark the letter that should come next in the series.

Number. This is the ability to work with figures and to handle simple quantitative problems rapidly and accurately. It is measured by asking subjects to check simple addition problems. The answer for each problem is provided, but some solutions are incorrect and the task of the subject is to check whether columns of figures have been added correctly or not.

Word Fluency. This ability is concerned with verbal recall involved in writing and talking easily. It differs from Verbal Meaning in its focus on the speed and ease with which words are used, rather than the degree of understanding of verbal concepts. The measurement task requires the subject to write as many words as possible beginning with the letter *s* during a 5-minute period.

Review of Earlier Work on the SLS

The Cross-sectional Sequence. The first cross-sectional inquiry, in 1956, suggested that for our oldest subjects (the mean age 67) relative performance remained best for Word Fluency and Number but was worst for men on Reasoning and for women on Space. Age difference gradients appeared to be steepest for Space and Reasoning and smallest for Number (see Fig. 3.2). Except for the youngest age/cohort, there was greater dispersion of the ability profile for women than for men. Another noteworthy finding was the differential attainment of peak performance levels. The peak for Reasoning occurred at age 25, whereas peak performance for Word Fluency was not attained until age 46; and for women, Number peaked only at age 53.

Patterns of ability at age 25 for our fourth wave (first tested in 1977) remained similar to those in our first study, but the sex difference in profile dispersion no longer prevailed. Extension of the cross-sectional data base into the 80s revealed somewhat different patterns at the oldest ages. The oldest male age/cohort in 1977 performed relatively best on Space and Reasoning and worst on Verbal Meaning. By contrast, the oldest women did best on Word Fluency and worst on Space and Verbal Meaning.

Cross-sectional age-difference data are not directly relevant to testing propositions about age-related change. Within the context of a cross-sectional sequence, however, they permit testing the proposition of whether patterns of age differences remain invariant over time. They also allow estimates of the magnitude of cohort differences and time-of-measurement effects. Our early analyses addressed the question of the invariance of age-difference patterns and concluded that there were statistically significant shifts, based on the findings of significant age-by-time interactions in time-sequential analyses and cohort-by-time interactions in time-sequential analyses and cohort-by-time interactions in cross-sequential analyses (cf. Schaie, Labouvie, & Buech, 1973).

Cross-sectional sequences also permit estimates of the magnitude of differences between successive cohorts and times of measurement. We can do so by averaging differences over successive measurement points for the different cohorts and likewise averaging differences across cohorts (at the same ages) over times of measurement. When this is done, it becomes clear from our data that there are systematic advances in cohort level for Space and Reasoning; a significant disadvantage is noted for the older cohorts up to that born in 1931. A similar pattern prevails for Verbal Meaning, although here only cohorts born in 1917 or earlier are significantly disadvantaged. Very different findings, however, occur for Number and Word Fluency. Number shows positive cohort changes up to about 1910. Then there is a plateau and a shift to successive lowering of performance level; the 1924 cohort exceeds both earlier and later born cohorts, and both oldest and youngest cohorts are now at a disadvantage compared to the

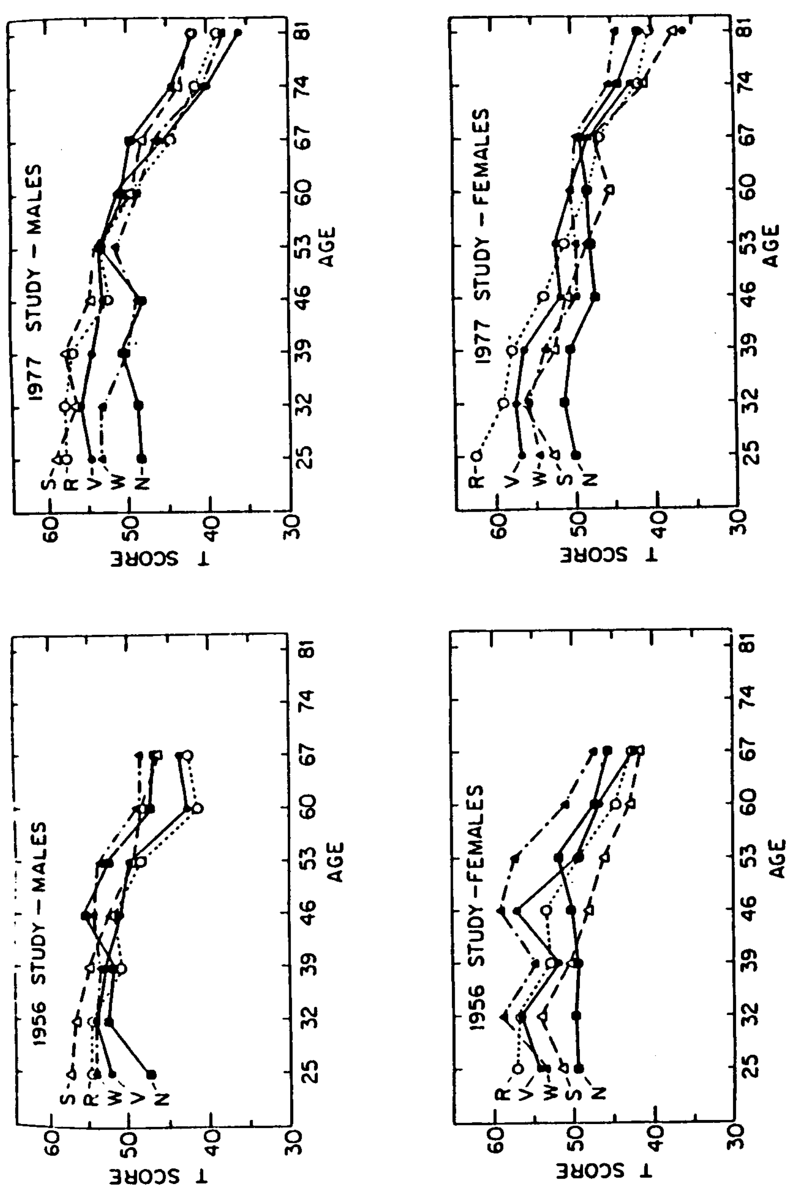

FIG. 3.2 Mean scores for the five PMA subtests for the 1956 and 1977 cross-sectional studies, by sex (from Schaie, 1983).

middle-aged! For Word Fluency we found successive lowering of cohort levels (the younger cohorts thus being disadvantaged) until 1938, but improvement for our two youngest cohorts.[5]

Time-of-measurement effects can be estimated from our data for three periods: (1) from 1956 to 1963 the only significant effect observed was a negative shift for Word Fluency; (2) there was a positive trend for all our variables from 1963 to 1970; (3) from 1970 to 1977, stability prevailed, except for a significant negative trend on Number.[6]

The Longitudinal Sequences. Early longitudinal analyses were concerned primarily with identifying ages at which statistically significant decline could be noted, as well as the determination of patterns of individual decline (cf. Schaie, 1979; Schaie & Labouvie-Vief, 1974; Schaie & Parham, 1977; Schaie & Strother, 1968).

Our findings immediately confirmed that the steep age decrement curves suggested by the cross-sectional data were an artifact of the confounded cohort effects. There was some question, however, whether the more gentle gradients and later points of significant decline that we estimated from our longitudinal data could have been attributable to longitudinal design threats such as practice or the substantial experimental attrition that is inevitable in longitudinal studies. Our comparisons of longitudinal and cross-sectional sequences indicate very few, if any, practice effects over the seven year intervals (see Schaie, 1983). We have determined, however, that retest survivors score higher at first test than individuals who drop out (see Fig. 3.3). Selective drop-out is particularly severe, however, only on the second test occasion. Although there is some age-by-dropout interaction, generally dropout seems to affect level of age gradients rather than their slope.

Substantive results from the longitudinal sequences can best be conveyed by estimates of cumulative age changes from within-subject data (cf. Schaie, 1983). Fig. 3.4 presents longitudinal estimates obtained by averaging over all 7-year longitudinal data for each 7-year age range, adjusted for cohort differences and time-of-measurement effects.

The average 7-year data consistently suggest the attainment of an adult plateau in the thirties, one that is generally maintained to age 60 (albeit with almost

[5]The complex issue of substantive interpretations of cohort and time-of-measurement effects has been discussed elsewhere (Schaie, 1984).

[6]Although cohort effects are always substantive in nature, time-of-measurement effects may simply represent systematic testing effects. That is, small but systemic variation in test administration and scoring procedures can slip into longitudinal studies even with the best standardization of methods. Likewise, although improbable for large samples it is nevertheless possible that our time-of-measurement differences could represent systematic sampling error, attributable to changes in the composition of the pool from which the successive samples were drawn.

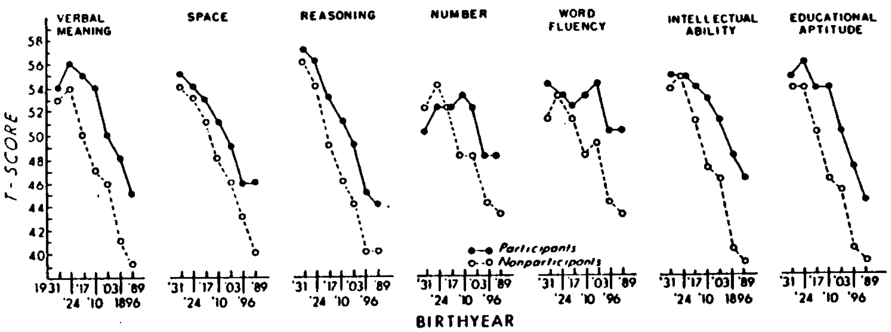

FIG. 3.3 Cohort differences for retest participants and dropouts (adapted from Schaie, Labouvie, & Barrett, 1973).

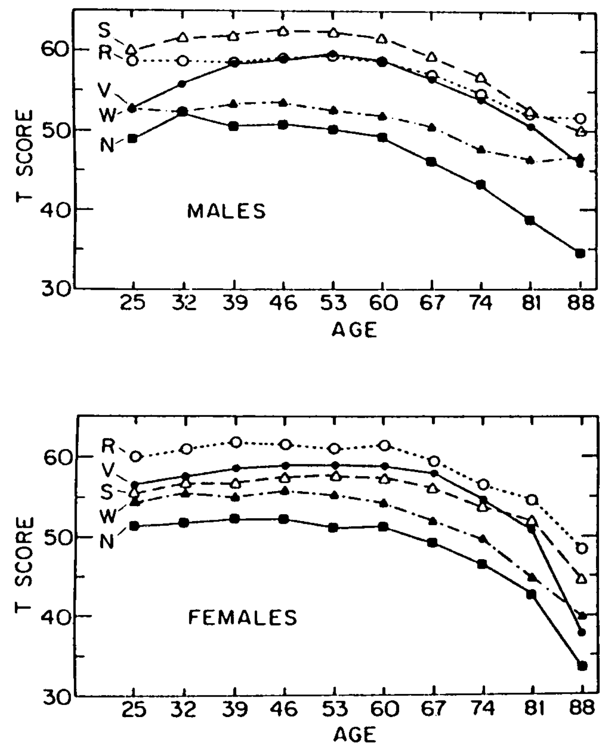

FIG. 3.4 Longitudinal estimates of age changes for men and women for the PMA subtests from pooled seven-year longtudinal sequences (from Schaie, 1983).

trivial decline for some abilities in the 50s). Thereafter, we note significant and accelerating average decrement that occurs at different rates by gender and ability. Thus, decline on Space seems to occur faster for women than for men, and the opposite holds for Word Fluency. However, the absolute magnitude of decline remains low (less than ½ S.D.) until the eighties are reached.

Cohort-Sequential MANOVA Analyses

Some of our early cross-sequential analyses (e.g., Schaie & Strother, 1968; Schaie & Labouvie-Vief, 1974) have been validly criticized because of two types of methodological flaws. First, the cross-sequential design, by crossing cohort with time-of-measurement, assumes that no age-associated effects are present in the data, thus impairing the usefulness of this approach for the precise estimation of age changes. Second, our early analyses compounded the problem by using unequal sampling intervals, that is, more cohort birth years than years of measurement. If differentiation of cohort and age effects is of primary interest, then the cohort-sequential design, by crossing age and cohort factors, is obviously more appropriate for the explicit estimation of age changes that are consistent across multiple cohort groups (cf. Schaie, 1977; Schaie & Hertzog, 1982). Such analyses were conducted recently for 7-year and 14-year cohort-sequential data sets (Schaie & Parham, 1977; Schaie & Hertzog, 1983) and are here described briefly.

Analysis Design. The data involved represent two longitudinal sequences that follow seven birth cohorts (partitioned into 7-year intervals) over a 14-year period. These sequences yield seven separate cohort-sequential data sets in which two adjacent birth cohorts are followed over the same 14-year age range. In addition, our cross-sequential sequences were similarly divided into directly comparable data sets. The age ranges studied were: 25 to 39 (for cohorts born in 1931 and 1938); 32 to 46 (cohorts born in 1924 and 1931); 39 to 53 (cohorts born in 1917 and 1924); 46 to 60 (cohorts born in 1910 and 1917); 53 to 67 (cohorts born in 1903 and 1910); 60 to 74 (cohorts born in 1896 and 1902); and 67 to 81 (cohorts born in 1889 and 1896).

The MANOVA paradigm used involved the method of hierarchical least squares (Bock, 1975). Linear and quadratic polynomial contrasts were used to represent age. Given the nonorthogonal designs, two models were computed: Age was fitted before cohort and vice versa. We used the partial sums of squares for hypothesis testing for age and cohort effects. Main effect and interaction parameter estimates for each factor were similarly adjusted.

Summary of Results. Significant effects for age and sex were found for all data sets in the cross-sectional sequences, and significant cohort effects for all

TABLE 3.1
Linear and Quadratic Age Effects and the Associated Proportion of Variance Statistics

		Linear effects									
		Verbal Meaning		*Space*		*Inductive Reasoning*		*Number*		*Word Fluency*	
Data set (mean age)	*Sequence*	*d*	ω^2	*d*	ω^2	*d*	ω^2	*d*	ω^2	*d*	ω^2
1	C	1.51	—	−1.33	—	−1.39	—	2.05	.01	−1.94*	.01
(25–39)	L	2.94**	.02	1.98*	—	1.30*	.01	1.27	—	2.24*	.01
2	C	−.92	—	−1.65	—	−2.51**	.01	−1.49	—	−3.46***	.02
(32–46)	L	1.76**	.01	1.45	—	.37	—	−.24	—	−.45	—
3	C	.63	—	−1.75	—	−1.27	—	.11	—	−1.26	—
(39–53)	L	.36	—	−.17	—	−.64	—	−.85*	—	−1.94	—
4	C	−1.24	—	−2.10**	.01	−1.97	—	−.53	—	−3.70**	.01
(46–60)	L	.84	—	−.04	—	−.10	—	−1.31*	.01	−2.83***	.01
5	C	−2.87*	.01	−2.39**	.01	−3.64***	.02	−2.33	—	−5.72***	.04
(53–67)	L	−1.74**	.01	−1.69*	.01	−1.55*	—	−1.90**	.01	−3.19***	.02
6	C	−5.06***	.05	−4.28***	.04	−3.32***	.03	−3.67***	.02	−7.14***	.09
(60–74)	L	−3.32**	.02	−2.76*	.01	−3.08***	.03	−2.16*	.01	−3.02***	.03
7	C	−5.83***	.07	−3.45***	.04	−2.73***	.02	−4.29***	.03	−4.30***	.03
(67–81)	L	−6.30***	.09	−5.48***	.08	−5.27***	.08	−5.33***	.07	−6.98***	.07

		Quadratic effects									
1	C	−.17	—	−1.94	—	−1.63	—	−6.03**	.02	1.61	—
(25–39)	L	−2.13	—	−1.94	—	−2.56	.01	−5.74**	.02	2.06	—
2	C	−.41	—	.65	—	.27	—	−2.73	—	3.14*	.01
(32–46)	L	−1.00	—	.04	—	−1.39	—	.42	—	2.56	—
3	C	.55	—	−1.40	—	.31	—	−1.69	—	1.87	—
(39–53)	L	−.95	—	−.43	—	1.56	—	−3.53***	.01	.10	—
4	C	1.31	—	−.57	—	2.58	—	1.38	—	3.60*	.01
(46–60)	L	−.88	—	−1.37	—	1.38	—	−.43	—	1.80	—
5	C	−4.62**	.01	−2.42*	.01	−3.27*	.01	−3.78*	.01	−2.10	—
(53–67)	L	−3.64**	.01	.05	—	−.88	—	−4.19***	.01	1.96	—
6	C	−1.76	—	.86	—	.70	—	−.57	—	3.29	—
(60–74)	L	.06	—	2.90	—	−2.07*	.01	−2.04	—	−.64	—
7	C	−.31	—	.95	—	.79	—	−.47	—	2.08	—
(67–81)	L	−4.55**	.01	1.23	—	.65	—	−3.56*	.01	1.60	—

Note. C = cross-sectional, L = longitudinal. Linear and quadratic effect contrasts, expressed in *T*-score units (.1 *SD*/unit). Negative linear contrasts indicate decline over the 14-year period; negative quadratic contrasts indicate downward concavity off the quadratic trend.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$. (From Schaie & Hertzog, 1983)

but the comparison of cohorts 1910 and 1917. Linear age effects occurred for all but the set ranging from age 39 to 53. All but the youngest and oldest set had significant age by cohort interactions. The two youngest data sets had a mixture of linear effects (some increasing, some decreasing). But, beginning with the data set originating at age 46, there was linear decline for at least some abilities. The cohort effects typically represented better performance for the most recently born, but this pattern was reversed for Word Fluency. Sex effects, consistent across all data sets, showed that men performed better on Space and Number, whereas women excelled on Word Fluency.

In the longitudinal sequences, significant age effects occurred throughout. These involved increment in performance until age 46, but also decline for some abilities during the age range from 46 to 60. Fewer cohort and sex effects were statistically reliable in the longitudinal sequences.

The results of the cohort-sequential analysis confirm that over some portions of the life span, there is indeed age-correlated decrement that is independent of generation differences. Given that these cohort-sequential estimates provide our best estimates of age-correlated change in PMA performance, we report the linear and quadratic parameter estimates for each PMA subtest in Table 3.1. The decline becomes evident after age 60 and (depending upon ability) ranges between ⅓ and ½ standard deviation over a 14-year interval. The present analyses also suggest small but statistically significant decline for some PMA subtests prior to age 60. Thus the cross-sectional sequence suggested a 2/10 standard deviation decline over the 14-year range from 46 to 60, and small declines were found in the data set ranging from age 53 to age 67. Once again, it should be stressed that the declines prior to age 60 are quite small and account for no more than 1% of the variance. However in the 60 to 80 year range, the declines in PMA performance become more pronounced for all abilities. Perhaps the most surprising finding is that the apparent differential patterns of age changes across PMA subtests observed in the cross-sectional data (see Fig. 3.2) are attenuated in these cohort-adjusted estimates. For example, although it is still the case that Verbal Meaning shows decline latest in the life span, the slope of the decremental function after age 60 is as steep as for other PMA subtests.

The cohort-sequential analysis further indicates that there are reliable cohort differences for Reasoning, Verbal Meaning and Space favoring more recently born cohorts. Major breaks between cohort pairs appeared to occur between the 1931 and 1938, the 1924 and 1917, and the 1910 through 1896 birth cohort pairs. These differences ranged between 2/10 and 1/2 standard deviations. Cohort effects favoring the older cohorts were found consistently for Word Fluency. These cohort-sequential cohort effects cannot be argued to be confounded age changes. Their existence confirms the concern that simple cross-sectional age differences grossly overestimate the magnitude of age change, especially for Reasoning and Space.

INDIVIDUAL DIFFERENCES IN INTELLECTUAL CHANGE

To this point we have reviewed the contributions of the SLS to addressing the questions of normative changes in level of psychometric intelligence over the adult life span. As we have seen, the SLS data, in agreement with other longitudinal studies on intelligence (e.g., Cunningham & Owens, 1983; Siegler, 1983), indicate reliable declines in mean test performance that begin to reach appreciable magnitude after age 60. We have also used the SLS data to address two other broad questions of critical importance in studies of adult intellectual development: (1) are there qualitative shifts in the nature of the psychometric tests, as indicated in different factor structures at different ages; (2) are there individual differences in age change in late adult life that covary with individual differences dimensions such as life style and physical health? The second question is obviously of primary importance in gerontological research—to scientists and lay persons alike! Who suffers the most decline, and why? Answers to these questions are a necessary precursor to addressing the question of central concern to all: what can we do to maximize our intellectual functioning in old age?

Methodological Issues in Assessing Factorial Invariance

The question of qualitative shifts in factor structure is in a sense a necessary first step to meaningful individual differences analyses with psychometric variables. Factor analysis of multiple indicator models is useful to determine whether the relationships between the multiple indicators and the factors may be taken to be invariant across multiple age (or cohort) groups or across age levels in longitudinal data (Baltes & Nesselroade, 1973; Mulaik, 1972; Nesselroade, 1970; Schaie & Hertzog, 1985). As discussed in some detail by Baltes and Nesselroade (1970, 1973), a lack of factorial invariance across developmental levels could be produced by changes in the measurement properties of the psychometric tests (which we shall refer to as *measurement equivalence*) or a change in the fundamental properties of the constructs themselves (termed *construct equivalence*). With respect to psychometric intelligence, a lack of measurement equivalence might indicate a shift in the relative importance of performance-related processes (e.g., perceptual analysis of form, selective attention) necessary for adequate performance on the tests but clearly distinct from the construct(s) the tests were originally intended to measure. Gerontologists have long concerned themselves with the measurement validity of psychometric tests in elderly populations (cf. Demming & Pressey, 1957; Schaie, 1977). On the other hand, a lack of construct equivalence across age levels might indicate a developmental shift in the organization of cognitive processes and their application to solving psychometric test items. To use Sternberg's (1980) terminology, shifts in the nature of metacomponents of intelligence test performance, such as task monitoring and strategy formation, might lead to changes in the covariance structure among the test scores. For example, an age-related shift toward a verbal propositional strategy for solving spatial orientation items might greatly increase the correlation between verbal fluency and spatial orientation in older populations. Although this specific question is best addressed with carefully designed experimental studies, empirical demonstration of a lack of factorial invariance would indicate that shifts in measurement equivalence, construct equivalence, or both are occurring during adult development—thus setting the stage for further work to isolate the nature of such changes.

Assessment of factorial invariance properties of multiple indicator models is therefore important, for only when factorial invariance is demonstrated can we assume that quantitative comparison of changes in developmental trajectories reflect changes in an isomorphic construct (Baltes & Nesselroade, 1970, 1973). Moreover, the meaning of differential change patterns (and their prediction by other individual difference variables) would be questionable if such differences reflected only shifts in the measurement properties of the tests in different subpopulations.

The extant factor analytic literature on adult age changes in intellectual factor structure has suggested a *dedifferentiation* phenomenon in which ability factors become increasingly intercorrelated to the point of a reduction in the number of factors needed to account for the correlations among intelligence tests (cf. Reinert, 1970). The critical demonstration of dedifferentiation in the absence of obvious reductions in the dimensionality of the factor space is seen in changes in the factor pattern (factor loading) matrix. The early literature is inconsistent—some studies found invariant factor patterns across age groups, but others have not (Reinert, 1970). This inconsistency may be attributed in part to problems associated with the analysis of multiple groups using separate exploratory factor analyses of correlation matrices from each group. As shown by Meredith (1964; see also Bloxom, 1972), under the process of selection from a population into subpopulations differing in mean levels of the factors (as is the case in cross-sectionally defined groups), only the unstandardized (raw score) factor loadings may be expected to be invariant over groups. In principle, selection will produce group differences in observed variables and factors; separate standardization in each group by dividing through by the standard deviations will therefore tend to obscure factorial invariance by producing artifactual differences in standardization factor loadings (see also Cunningham, 1978; Hertzog & Carter, 1982; Jöreskog, 1971). Recent cross-sectional work by Cunningham (1980, 1981), using appropriate confirmatory factor analysis procedures, has suggested invariance in unstandardized factor loadings for certain primary abilities across multiple age groups.

We have investigated issues of factorial invariance in our longitudinal data

by application of longitudinal models developed by Jöreskog and Sörbom (1977, 1980). These models avoid the problems of standardization described previously and are uniquely suited to testing multiple hypotheses of factorial invariance in longitudinal sequences (Jöreskog, 1979; see also Hertzog & Schaie, 1982, 1984). Specifically, one can use the longitudinal models to obtain: (1) tests of age-related factorial invariance within groups measured longitudinally; (2) evidence regarding the stability of interindividual differences in intraindividual change, as indicated by the magnitude of factor covariances over time; (3) evidence of changes in factor interrelationships, as defined by differences over time in the within-occasion covariances among different factors; (4) separation of stability of IEVs in IACs for subtest-specific components from stability of IEVs in IACs in common factor scores.

Longitudinal Factor Analysis of the Primary Ability Tests[7]

Our interest in the issues of factorial invariance and stability of IEVs in IACs, measured at the level of the second order factors, led to an analysis of the stability of the second order general intelligence factor that can be derived from the 1948 PMA subtests. Thurstone and Thurstone (1941) had originally demonstrated that the factor correlations among the six primary abilities they identified (which included a memory factor dropped from the 1948 PMA battery) could be factored at the second order level to obtain a general factor they identified as being analogous to Spearman's *g*. Our preliminary analysis of the PMA subtests, in conjunction with measures from Schaie's Test of Behavioral Rigidity (Schaie, 1958), indicated convincingly that a single second order factor gave a plausible representation of the covariances among the PMA measures. The subsequent analyses were therefore devoted to examining the characteristics of this second order factor in longitudinal data.

In order to achieve adequate sample sizes for factor analysis, the data from the two 14-year longitudinal sequences were reorganized under the assumption of no cohort differences in factor structure to yield three age groups ("young," "middle-aged," and "old" at time of first test) who were all followed over the 14-year period (see Table 3.2). The longitudinal factor analysis model for the data is shown in Fig. 3.5: the general PMA factor was modeled at each longitudinal occasion, the factor covariance matrix was left unconstrained, and the residual covariance matrix contained unique (specific) variances on the diagonal and autocorrelated residuals for the specific PMA subtests off the diagonal, a

[7] The analyses reported here are part of an unpublished doctoral dissertation by Hertzog (1979) and a reanalysis of the same data reported by Hertzog and Schaie (1984). We cannot summarize in the available space many of the salient features of the analyses reported in those references, particularly details regarding the formal specification of the structural regression models.

TABLE 3.2
Reparameterized Sequential Sample for Multiple Group Analysis

	Sample	*Cohort (mean birth year)*	*Age (means)*	*N*
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

critical feature of most longitudinal models (Corballis, 1973; Sörbom, 1975).[8]

The first order of business was to determine whether the *g* factor had the same relation to PMA subtests across the adult lifespan. To do so, we tested the hypothesis of invariance in the unstandardized factor pattern matrix both between groups and across occasions. These tests were obtained by taking differences in the likelihood ratio x^2 tests of fit between (a) the models constraining corresponding factor loadings to be equal, either between groups or across occasions; and (b) the model with the factor loading specified to have the same free parameters but with no equality constraints placed upon them (see Alwin & Jackson, 1981; and Jöreskog, 1971, for an explication of the logic of such hypothesis tests). The results of these tests were unequivocal—there was no loss of fit by requiring the factor loadings to be invariant both across occasions and between groups (see Hertzog & Schaie, 1984, for details). We therefore concluded that the general factor is qualitatively invariant across the adult life span.

There were, however, reliable group differences and occasion-specific differences in factor variances and covariances. There was a substantial increase in G variance from the second to the third occasion of measurement for the

[8]In fact, an initial factor analysis model omitting these residuals produced a badly perturbed solution, with factor covariances exceeding the maximum possible bound for the estimated factor variances (see Hertzog, 1979).

	G_1	G_2	G_3
V_1	λ_1	0	0
S_1	λ_2	0	0
R_1	λ_3	0	0
N_1	λ_4	0	0
W_1	λ_5	0	0
V_2	0	λ_6	0
S_2	0	λ_7	0
R_2	0	λ_8	0
N_2	0	λ_9	0
W_2	0	λ_{10}	0
V_3	0	0	λ_{11}
S_3	0	0	λ_{12}
R_3	0	0	λ_{13}
N_3	0	0	λ_{14}
W_3	0	0	λ_{15}

FIG. 3.5 Factor loadings (Λ) for longitudinal factor analysis models occasion-specific Λ

oldest group and a parallel but smaller change in the middle aged group, but the G variance actually decreased from the first through the third occasion in the young group. The factor covariances were uniformly high for all groups. When standardized, the correlations among the factors were generally at the .9 level or higher.

Given the longitudinal nature of the data, these factor covariance parameters have special implications for the interpretation of individual differences in developmental patterns in these age groups. The high factor covariances indicate an impressive degree of stability in interindividual differences under the presence of intraindividual change; that is, individuals seem to be preserving to a remarkable degree their relative orderings about the *g* factor mean over the 14-year interval in all age groups. This finding is tempered somewhat by the increasing variances in *g* in the two older groups. The joint pattern of *g* variances and covariances suggests that, although there is some divergence of individuals in terms of magnitude of change in *g* over time, this divergence is sufficiently weak to fail to alter dramatically the pattern of IEVs observed at the first occasion. The data are therefore consistent with the interpretation that changes in *g* are relatively parallel across individuals.

This interpretation must be considered in light of the likely effects of experimental mortality. Attrition of low *g* subjects might tend to increase the stability observed in the data, inasmuch as persons undergoing nonnormative decline secondary to physiological or psychological factors are relatively more likely to be lost from the panel study. In short, we are probably looking at a positively selected subpopulation in which greater relative stability in interindividual change should be expected than in the general population. Nevertheless, the magnitude of the stability of *g* is impressive.

Is this stability present at the *g* factor level, but not at the primary ability level? One could argue that individual differences in change are minimal for general intelligence but more likely in specific abilities (e.g., Spatial Orientation, Verbal Comprehension, where individual differences in life experiences of special relevance to maintenance of specific abilities might wield more importance. Although we cannot examine this hypothesis by computing factor covariances for the primary abilities, this issue is in part addressed by the covariances among the residual (specific) components across occasions. In the occasion-specific model, all these residual covariances were statistically reliable, indicating considerable cross-occasion consistency in interindividual differences at the level of the primary abilities as well. We confirmed this impression with a model that added five test-specific factors to the occasion-specific factors for the groups. The test-specific factors posited a Verbal Comprehension factor, a Spatial Orientation factor, etc. These factors were formed by forcing each measure to load on its own test-specific factor in addition to *g* at a given occasion. The test-specific factors were forced to be orthogonal to the occasion-specific G factors by fixing their covariances to zero.

We were able to form salient test-specific factors for all five PMA subtests in this model. The fit of the model was good ($x^2 = 287.68$, df $= 248$, $p = .04$). Table 3.3 gives the relevant information regarding the parameter estimates for the model. In addition to the significant loadings, the values of the residual matrix show two values for specific variances: (a) those of the occasion-specific

TABLE 3.3

	OCC/TST	*OCC*	*LY**	*TST*	*TE1*	*TE2*	*TE3*
V1	1.67(.11)			1.03(.16)	.12(.45)	.13(.40)	.28(.37)
S1	.95(.09)			1.00(.08)	.14(.86)	.37(.85)	.40(.75)
R1	1*			.752(.16)	.26(.45)	.18(.48)	.21(.32)
N1	1.46(.11)			1.01(.08)	.13(.72)	.11(.70)	.14(.58)
W1	1.34(.12)			.67(.10)	.47(.80)	.30(.81)	.23(.74)
V2		1.66		1*	.08(.47)	.13(.44)	.24(.43)
S2		.95		1*	.21(.90)	.14(.82)	.43(.72)
R2		1*		1*	.25(.57)	.16(.42)	.13(.34)
N2		1.46		1*	.20(.78)	.12(.72)	.14(.60)
W2		1.34		1*	.17(.86)	.20(.76)	.21(.72)
V3			.77	.96(.15)	.17(.53)	.18(.38)	.17(.35)
S3			.44	.96(.08)	.27(.91)	.42(.81)	.26(.64)
R3			.78	.99(.18)	.23(.54)	.15(.40)	.11(.22)
N3			.59	.96(.08)	.10(.76)	.12(.64)	.15(.47)
W3			.48	.98(.13)	.26(.88)	.22(.73)	.21(.65)

Note: First column gives occasion specific LY coefficients in model with test specific factors. Second column are the same coefficient estimated with the occasion specific model only. The third column (LY*) represents scaling which preserves occasion specific invariance. Three columns for TE show proportion of unique variance for groups 1, 2, 3, (young, middle aged, old) for the occ/tst model; the proportions of unique variance in the occasion specific model are given in parentheses.

model; and (b) those of the model with both test-specific and occasion-specific factors. These values reflect the proportion of variance in each subtest that both is specific to the subtest and also covaries across occasions. In the case of Verbal Meaning at occasion 1, we see that 45% of the variance was unique in the occasion specific factor model, but only 12% is unaccounted for when the test-specific factor is added. A scan of these residual variances in Table 3.3 reveals two important facts about the stability of interindividual differences in these data. First, for all variables, there is substantial stability that is specific to the primary ability measures (i.e., independent of *g*). Second, the residual variances in the test-specific model are very close to the values one would predict from the published reliabilities of the tests. For instance, the 1948 PMA manual reported an estimated reliability of .92 for Verbal Meaning. Hence, we may conclude that there is substantial stability in individual differences at the primary ability level as well.

This stability of interindividual differences should not be confused with changes in the PMA means over time. The data suggest that there are age-related changes in mean performance levels and that individual differences are relatively consistent in spite of such changes. This point was underscored when a simultaneous analysis of means and covariance structures was performed on the longitudinal data. In spite of the high stability of individual differences in *g* in the older group, their means show progressive declines in performance. On the other hand, the data for the middle-aged group may be modeled as stable in both level and covariance structure over time. Thus, the results seem to indicate a pattern of relative stability of performance levels during the decades of the 40s and 50s, but a shift to a pattern of performance decline following age 60. This shift seems to be relatively pervasive with respect to both subtests and persons.

In summary, our longitudinal factor analyses indicate two important conclusions regarding individual differences in intellectual change in adulthood: (1) the changes are not of the nature to alter the fundamental covariance structure of the abilities, since invariant raw score factor pattern matrices may be isolated at each age level (assessed both across the age groups and longitudinal occasions for the same subjects); and (2) there is an impressive degree of stability of interindividual differences in PMA performance across the adult life span. Although our results probably overestimate the magnitude of such stability due to positive bias inherent in long-term longitudinal samples, it is interesting to observe such stability even if it is restricted to a subpopulation of older adults.

Predictors of Individual Differences in Intellectual Change

Despite the high degree of stability in our longitudinal samples, we have identified certain variables associated with differential change in the latter portions of the life span. One set of variables includes various dimensions of physical health. We have been able to verify empirically what may seem to some to be the obvious, namely, that good physical health in middle and old age is associated with continued participation in our longitudinal samples and with maintenance of levels of intellectual functioning (Hertzog, Schaie, & Gribbin, 1978; Parham, Gribbin, Hertzog, & Schaie, 1975). The study by Hertzog, et al. (1978) provided the clearest evidence of a disease-related attrition process that results in overestimation of the stability of intelligence. It reported that presence of severe forms of cardiovascular disease (CVD) was significantly associated with loss of subjects from the first longitudinal sequence. This attrition effect was sufficient to reduce relationships between CVD and intellectual change to nonsignificant levels in the surviving subjects. However, when data for all subjects who had participated in the first two occasions of measurement were analyzed, we found that certain forms of CVD (cerebrovascular disease and coronary atherosclerosis) were related to greater decremental changes during the 53–74 age ranges than for same-cohort healthy comparison groups. The proportions of variance accounted for are relatively small, however. Thus, the correlations between disease and magnitude of intellectual change reported by others (e.g., Wilkie & Eisdorfer, 1971) can be observed in SLS data. It does appear, however, that nonmorbid levels of these diseases are not associated with differential loss, as judged from the data from the 14-year longitudinal sample.

The question remains as to the causal nature of this relationship. Although most individuals would be inclined to interpret such data as indicating a physiologically mediated deficit in brain function (e.g., through reduced oxygenation of brain tissue), it is also the case that severe health problems are associated with reduction in activity patterns, depression, and other variables that could mediate the disease/intelligence relationship. No one would question whether cerebral ischemic events are likely to produce cognitive deficits. However, one could argue that relatively mild forms of CVD lead to life-style changes that, in turn, more directly affect cognitive behavior. Further studies of disease relationships to cognitive behavior by Stone (1980) may bear indirectly on this issue. She conducted structural equations analyses on several of the longitudinal samples. Disease categories investigated included circulatory disorders, digestive disorders, genitourinary problems, diseases of the musculoskeletal system, and diseases of the nervous system and sense organs, among others. No relationship between health and PMA performance were found in the young to middle-age ranges. However, in the older subjects, circulatory disorders, neoplasms, and musculoskeletal disorders were found to be significant predictors of intellectual change over 7-year intervals. The proportions of variance accounted for were relatively small, ranging from 1.6% for neoplasms to 4.4% for CVD. Again, interindividual differences in intelligence were very stable, accounting for about 85% of the interindividual differences. However, it is more difficult to posit direct physiological mechanisms for musculoskeletal disorders and, to a lesser extent, neoplasms (not of central nervous system origin) on brain functioning. An alternative is to assume that severe arthritis and other musculoskeletal diseases

reduce the ambulation of elderly persons and may have detrimental effects on behaviors that maintain intellectual functioning.

We have also examined other factors associated with changes in intelligence, including patterns of life activities (life style), personality, and socioeconomic status (SES). Our preliminary findings are reported in more detail elsewhere (and summarized by Schaie, 1983). Hertzog (1979) found a latent SES variable to have strong levels of prediction to *g* in a structural regression model, but it did not predict intellectual change over time. Gribbin, Schaie, and Parham (1980) looked more closely at life style differences that might correlate with differential change and found some relationships. That study used cluster analysis of a life activities questionnaire to yield eight item clusters representing: (1) subjective dissatisfaction with life status; (2) SES; (3) noisy environment; (4) family dissolution; (5) disengagement from interaction with the environment; (6) semipassive engagement with the environment; (7) maintenance of acculturation; (8) female homemaker characteristics. Seven-year changes in PMA performance did correlate with some life style clusters. The analysis suggested that, in general, personal characteristics of disengagement and family dissolution are associated with greater cognitive decrement, whereas dissatisfaction with life status was associated with maintenance or improvement in functioning.

Gribbin et al. (1980) also examined the relationship between person clusters defined by life style variables and intellectual change. Four major person types emerged: Type 1 persons were mostly males of average social status and average level of acculturation who lived with intact families and who were quite engaged with life (although they voiced strong dissatisfaction with their life status). Type 2 persons had high social status and high life satisfaction, intact families, and above average levels of maintenance of acculturation. Type 3 persons were mostly women of average social status who were homemakers with intact family situations and average life satisfaction. Type 4 persons were almost all female, but on average older than Type 3 females. They had low social status, were relatively dissatisfied, and were lowest on reported activities, whether active or passive in nature. Table 3.4 lists the average change scores for the PMA subtests

TABLE 3.4
Intellectual Change over 14 Years by Lifestyle Types

	Type 1	*Type 2*	*Type 3*	*Type 4*	*All types*
Verbal Meaning	+3.6	+1.4	−.2	−4.4	+.4
Space	+1.0	−1.1	+1.5	−.1	+.2
Reasoning	−.2	+1.1	−1.9	−3.1	−.7
Number	+1.0	+.8	−.9	−1.6	
Word Fluency	−3.5	−4.2	−6.2	−5.6	−4.8
Intellectual Ability	+.5	−.2	−1.9	−4.0	−1.1
Educational Aptitude	+2.8	+1.3	−.6	−4.3	+.1

*Positive values denote incremental change; negative values denote decremental change.

over a 14-year period for each life style type. It is apparent that the first two types generally show maintenance or gain, whereas the most disengaged type, the widowed homemakers, show clear evidence of decrement.

NEW DIRECTIONS FOR THE SLS

During the past decade, there has been an increased focus in the study of adult intelligence on issues related to optimal or adaptive intellectual functioning. Researchers have attempted to address this issue in at least three major ways. First, there has been the search for new models of intellectual functioning uniquely characterizing cognition in adulthood. A second major thrust has focused on the modifiability of adult intellectual performance through cognitive training procedures. And third, major efforts of a microanalytic nature have sought to examine, within the information processing approach, components of intellectual performance that do or do not exhibit age differences. This section of the chapter describes some new departures for the SLS that move into such new directions.

Training Studies and the SLS: Remediation versus Improvement[9]

Even well-functioning older adults can be disadvantaged in at least two different ways: First, some age-related decline may occur through disuse, whether by personal choice or environmental restrictions; second, other older adults may be disadvantaged because of rapid sociocultural and technological change. Prior cross-sectional cognitive training research has strongly suggested that modifiability of older adults' performance on a number of intelligence dimensions (see Willis, 1985, and Willis & Schaie, 1981 for reviews of training research). However, the cross-sectional nature of this research made it impossible to examine one of the most fundamental questions. That is, to what extent did training procedures result in remediation of age-related decline versus the acquisition of new performance levels in subjects experiencing no decline? This issue must be addressed within the context of a longitudinal study, such that training improvement can be assessed for subjects with known intellectual histories. Research currently in progress seeks to obtain requisite data.

Training Design. Subjects from the five oldest cohorts (age 60+) of the Seattle Longitudinal Study were classified, on the basis of prior longitudinal data and pretest scores, as having either remained stable or shown significant decline in performance on two of Thurstone's Primary Mental Abilities, Spatial Orientation and Inductive Reasoning. Significant decline was defined as equal to or greater than one standard error of measurement. Subjects showing decline on one of the two abilities were assigned to training in the declining ability. Subjects who either remained stable on both abilities or declined on both abilities were randomly assigned to training in one of the abilities.

The study involves a pretest-posttest control group design. Subjects are pre- and posttested on an extensive battery of primary ability measures. Training subjects receive five one-hour training sessions in their homes. Training procedures focus on strategies and skills shown in previous research to facilitate performance on the target ability. Approximately 200 subjects are being trained, 100 on Spatial Orientation and 100 on Inductive Reasoning. Approximately 200 additional control subjects are also being tested. Two-thirds of the sample is female; ⅓ is male. For the ⅓ of the sample that has completed the study thus far, mean age of subjects is 73 years (Range = 62–94).

Preliminary Results. A primary concern in this new venture is to determine the proportion of subjects showing significant training improvement and to differentiate between returning to previous performance levels subjects who had declined and raising the performance of those who had remained stable over time.

Our first question is: what percentage of subjects showed significant training gains? Training gains are defined as equal to or greater than one standard error of measurement. It appears that training procedures are particularly effective for subjects who exhibited prior decline. The greater gains of the decliners, moreover, cannot be attributed to regression effects, as their decline had been defined over multiple measurement occasions (Nesselroade, Stigler, & Baltes, 1980). Of the decline subjects receiving Space training, 81% of the women and 58% of the men showed significant training improvement at post-test. Training resulted in remediation to subjects' 1970 pre-decline score level for 71% of the women and ⅓ of the men. Of decline subjects receiving training on Inductive Reasoning, 63% of the men and 56% of the women showed significant training improvement at post-test. Remediation of performance to their 1970 score level occurred for 50% of the men and 40% of the women.

Significant training gains were achieved by fewer of the subjects exhibiting no prior decline in intellectual functioning on the target ability. One-third of the stable women and 20% of the stable men trained on Space showed significant training improvement. For Reasoning, ⅓ of the stable men and 14% of the stable women showed significant training improvement at posttest.

Training improvement was compared also with performance of a control group receiving only the pretests and posttests. Training gain on Space for both stable and decline females was significantly greater than gain for control females.

[9]This brief review of our attempts to reverse intellectual decline in advanced age is a condensation of a more extensive presentation by Willis and Schaie (1983).

Females exhibiting prior decline showed the largest training effect. However, training gain for men was not statistically greater than for the control men.

These findings were based on posttest performance under standard time conditions. Posttest performance under relaxed-time conditions (double the standard time) was also assessed, and a second gain score (relaxed-time posttest score minus standard time posttest score) was computed. It was reasoned that subjects performing at a high level prior to training might require additional time at posttest in order to demonstrate training effects. The results of this analysis show that gain from times to relaxed time posttest scores was significantly greater for stable training subjects than for stable control subjects.

On Inductive Reasoning, subjects showing prior decline demonstrated significant training effects when compared with stable training subjects. These gains were significantly greater than those found for the control group receiving only pretests and posttests. Post hoc analyses indicated further that training effects were greater for the training males than for decline training females.

In summary, these preliminary data from subjects with known intellectual histories support the promising findings of previous cross-sectional training research and further clarify the nature of the training effects achieved. More important, strong evidence is provided that reliably measured age decrement can be selectively reversed for a substantial proportion of older subjects. Indeed, quite a few subjects can be returned to the level of functioning shown by them prior to decline some 14 years earlier. In addition, under relaxed time conditions, it is possible to improve the accuracy level of those elderly subjects who have not shown any intellectual decline. It should be noted that a major focus of training research conducted within a longitudinal design is on intraindividual change. That is, the critical comparison is to subjects' own prior performance level. These findings seem to lend support to "disuse" theories of cognitive decline and raise further questions about the universality of irreversible decline in intellectual functioning with advanced age. Given the limited sample sizes, representing only ⅓ of the final sample, these findings must be interpreted with caution.

Information Processing Models for PMA Performance

It is by now well recognized among cognitive psychologists that performance on psychometric intelligence tests can usefully be modeled in terms of the elementary cognitive processes required for test item solutions (e.g., Carroll, 1976, 1980; Carroll & Maxwell, 1979; Hunt, 1978; Pellegrino & Glaser, 1980; Sternberg, 1977, 1980). A number of studies have successfully used manipulation of item characteristics in reaction time (RT) paradigms designed to isolate individual differences in the use, speed, and successful completion of constituent information-processing steps leading toward responses on several types of psychometric test items, including spatial orientation (Egan, 1981; Lohman & Kyllonen, 1983;

Pellegrino & Kail, 1982), inductive reasoning (Pellegrino & Glaser, 1980; Kotovsky & Simon, 1973), and verbal comprehension (Hunt, 1978; Sternberg & Powell, 1983). A "componential" model of intelligent thought enables characterization of individual differences in intelligence in terms of: (1) individual differences in the use of qualitatively different strategies for task solution; (2) individual differences in the speed and accuracy of component step execution required by the strategy selected.

Such approaches have only recently been utilized to examine individual differences in the contribution of different cognitive processes in shaping adult intellectual development (e.g., Berg, Hertzog, & Hunt, 1982; Hunt, 1983). This state of affairs is rather surprising. Gerontologists usually interpret age changes in intelligence test scores as loss of ability in the construct domain nominally measured by the test. If spatial test scores decline, the inference is that the ability to imagine a rotation of form in space has declined. From a componential perspective, such a conclusion is at least premature, if not actually invalid. Inasmuch as psychometric test items require the concatenation of multiple processing steps, failure to achieve a correct solution in a fixed amount of time may be due to failure at any step along the way, selection of an inefficient or incorrect strategy, or slowed information-processing speeds.

Individual differences in the speed of component execution may determine adult age differences in psychometric test performance to a greater degree than has been widely understood. It is well known that primary biological aging causes slowing in the speed of virtually all nervous system functions (e.g., Birren, 1974; Birren, Woods, & Williams, 1979). Several theorists have suggested that many of the age differences observed in certain cognitive tasks may be inflated, if not completely determined, by age-related slowing in memory retrieval speed and other cognitive mechanisms (e.g., Salthouse, 1980; Waugh & Barr, 1980). Although we doubt that age-related cognitive slowing is the only determinant of changes in psychometric test performance, it stands to reason that some of the changes in PMA performance are a function of a slowing of processes needed for intelligent thought.

Under these circumstances, it would not be surprising to find that age-related changes in PMA performance, particularly in the 40–65 age range, were primarily a function of slowed information-processing speeds rather than loss of the basic ability to think intelligently per se. Preliminary evidence in favor of this hypothesis may be found in a study by Berg et al. (1982), who were primarily concerned with examining age differences in the mental rotations (MR) task. As is well known, the mental rotations paradigm developed by Shepard and colleagues (cf. Cooper & Shepard, 1978) has been used to assess the speed of an internally imaged, mental rotation of geometric stimuli. The paradigm requires subjects to make a YES/NO identity recognition of two figures of differing angular orientation. Mirror image figures are used as distractors. The MR task has also been used for analyzing individual differences in spatial orientation (e.g., Kail, Carter,

& Pellegrino, 1979; Lansman, Donaldson, Hunt & Yantis, 1982). Indeed, Kail et al. (1979) used the Thurstone's PMA stimuli in their mental rotations task (see also Pellegrino & Kail, 1982). The major distinctions between the PMA Space test and the MR task are the differences in method (experimental apparatus versus psychometric test) and format (single paired comparisons of figures in the MR task versus a series rows of items in the PMA, in which each row comprises six comparisons to a standard figure).

Berg et al. (1982) studied mental rotations RT performance in four different age groups and found reliable age differences in the slope and intercept parameters of the linear RT function relating response times on correct trials to degrees of angular disparity between two figures to be compared. Similar results have been reported for other stimuli (e.g., Cerella, Poon, & Fozard, 1981; Gaylord & Marsh, 1975). Of particular interest was the fact that MR slope and MR intercept had significant correlations (r's in the .5 to .7 range) with scores on the PMA Space test in the older groups. The correlations were higher for older groups and specific to spatial ability (correlations with Nelson-Denny vocabulary scores were not statistically reliable). Thus, the speed with which one rotates spatial images and makes an accurate figural discrimination correlates significantly with overall PMA Space test scores for older subjects. Moreover, the predictive value of the speed of information processing on correct trials is more highly related to older persons' PMA performance than it is to performance in younger groups.

In our view, such findings reinforce the notion that careful studies of the speed and accuracy of information-processing steps involved in psychometric test performance could contribute greatly to our understanding of age-related changes in intelligence. Although studies in the gerontological literature have shown age differences in psychometric performance under untimed administration of the tests (e.g., Schaie, Rosenthal, & Perlman, 1953), we do not know the extent to which construct-relevant speed components such as mental rotation speed, determine the changes in overall test scores we have been charting in the SLS. Furthermore, we have yet to assess whether age-related changes, when they do occur, may be attributable to changes in certain processes rather than to others, whether the loss of efficiency in information-processing functions follows some prescribed order, or whether certain patterns of change are generally associated with specific individual characteristics such as cardiovascular disease, alcohol abuse, or other aspects of adult behavior.

We are currently preparing to begin a cross-sectional study to measure the correlations between performance on two PMA subtests, Verbal Meaning and Space, and information-processing speed on conceptually relevant tasks. Although it is too early for us even to speculate on likely outcomes, a brief discussion of the research design and planned analysis will illustrate the potential value of our approach. We intend to collect data on more than 200 adults in the 45–75 age range. All individuals will first be tested in psychometric test sessions designed to measure several of the primary abilities identified by Thurstone and the ETS

Reference Kit (Ekstrom, French & Harman, 1976). We will measure the adults' performance on several RT tasks, including: (1) choice RT; (2) three tasks measuring semantic memory access speed (Hunt, Davidson, & Lansman, 1981); (3) a mental rotations task.

Figure 3.6 illustrates a structural regression model that can be used to examine the convergent and discriminant validity of the information-processing tasks for prediction of the Spatial Orientation and Verbal Comprehension factors from the psychometric battery (see Embretson, 1983, for a further justification of this type of approach). The model may be used to assess whether there is a general information-processing speed factor that correlates with individual differences

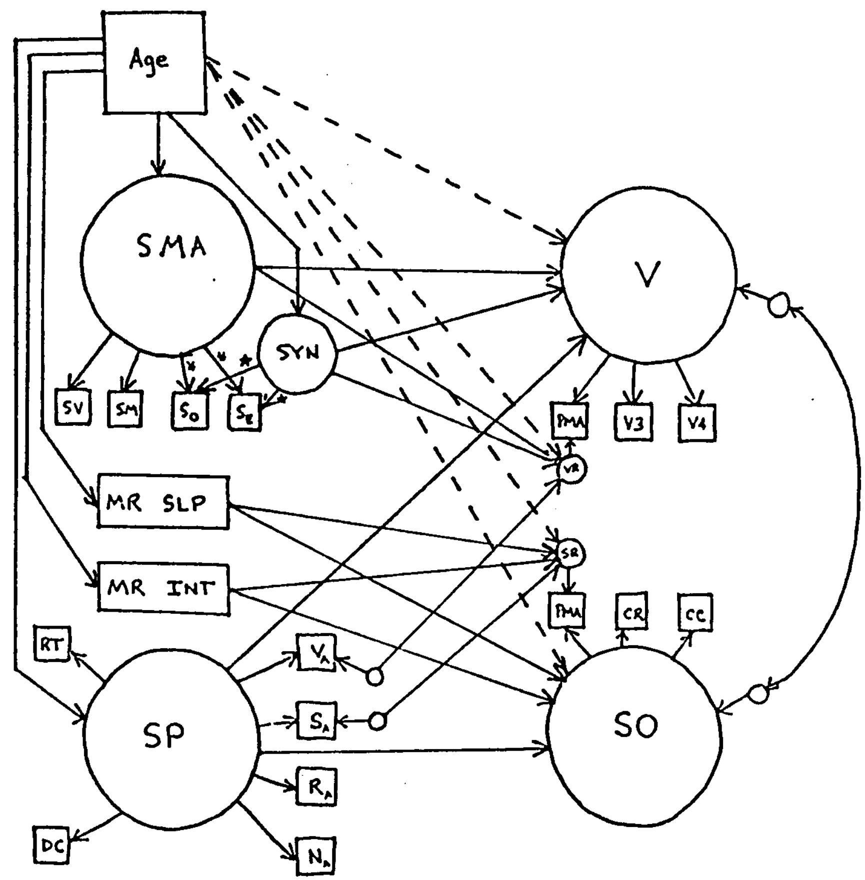

FIG. 3.6 Structural model regressing Verbal Comprehension and Spatial Orientation factors on information processing dimensions and chronological age.

in psychometric performance, or alternatively, whether the information-processing speed/psychometric tasks relationships differ according to the type of information being analyzed (spatial versus verbal). Moreover, we can use the regression model to determine whether age differences in these factors are statistically independent of age differences in the latency and accuracy data from the information-processing tasks. Although the study is limited by its adoption of a cognitive-correlates approach and by analysis of cross-sectional age differences in intelligence, it represents a useful first step toward a more complete information-processing analysis of individual differences in age changes on psychometric task performance. Our ultimate goal is the development of a set of componential tasks for PMA performance that could be incorporated into a longitudinal design like that of the SLS.

SUMMARY AND CONCLUSIONS

When the SLS began nearly 30 years ago, the predominant theoretical model in cognitive gerontology was one of massive and irreversible loss of cognitive capacity. Today it is widely recognized that such a model is simply too extreme. In the domain of psychometric intelligence, it is now understood that the previously dominant cross-sectional design overestimated the magnitude of age differences in most abilities. The data from the SLS suggests strongly that the primary abilities measured by Thurstone's PMA test do not show practically meaningful declines until after age 60. As we have seen, this conclusion, stated in forceful terms by Schaie and Strother (1968), has been upheld in subsequent studies from the SLS using a more complex sequential sampling design and more sophisticated statistical methods.

In our view, it is time for cognitive gerontologists to move away from studies focusing exclusively on identifying the point in the life span where mean levels of cognitive decline may be observed. Thus, our recent work was focused more directly on multivariate models of intellectual change, prediction of individual differences in change, and attempts to remediate previously observed decrement through training programs. The longitudinal factor analyses we reported provide important evidence regarding the factorial invariance of a second-order general intelligence factor from the PMA across the adult life span. This finding seems to rule out, at least at the level of general intelligence, any fundamental shift in the measurement properties of the PMA in older populations. Our longitudinal analyses also show a surprising degree of stability of interindividual differences in 14-year longitudinal samples, both for a second order *g* factor from the PMA and for the primary ability subtests themselves. As we have seen, there is reason to believe that this relative stability may be more likely for positively selected subpopulations. Nevertheless, the data indicate that groups positively selected for longevity and good health are characterized by stability of intellectual functioning (at least as measured by psychometric test performance) through middle age. However, roughly the same pattern of mean changes may also be observed in cross-sectional sequences, which are not prone to the kinds of testing and experimental mortality problems as the longitudinal sequences.

Physiological problems, notably cardiovascular and cerebrovascular disease, are correlated with differential decline and attrition from the longitudinal panel. Sequential studies specifically designed to measure the cognitive impact of such age-correlated diseases should undoubtedly use a shorter retest interval than the 7-year interval employed by the SLS. Other studies from the SLS suggest that individual differences in activity patterns in old age may be correlated with degree of decline. We found some indication that elderly, inactive, and isolated women—the "disengaged homemakers"—typically exhibited steep intellectual decline toward the end of the life span. Our data do not permit unambiguous causal inferences regarding the nature of the life style/disease/intelligence relationships, so some caution should be exercised. In the case of cardiovascular disease, for example, it is not certain that any age-correlated decline is physiologically mediated (i.e., due to alterations in oxygen supply to brain tissue) or psychologically mediated (restrictions of activity due to illness leading to lowered morale, reduced mental as well as physical exercise, and so on).

Our current efforts have involved (a) an investigation of the magnitude of benefit of training on psychometric test performance and (b) development of information processing approaches that will enable a more elaborate description of age changes in psychometric test performance. The training study has suggested that the benefits of training on psychometric tests are largest for those with previous declines in PMA performance. The information-processing analyses are founded on the premise that the transition to a decline pattern in PMA performance between ages 60 and 70 may be mediated primarily by age-related cognitive slowing in intellectual thought processes. If so, then the similarity in patterns of intellectual change in late life may be in part produced by the "masking" of individual differences in developmental change in capacity by the increasingly important role of a relatively ubiquitous slowing process. In any case, it is our belief that such models afford a better opportunity to uncover the (undoubtedly multiple) causes of intellectual change, because they will allow us to identify multiple alternative sources of poor performance that might otherwise be lumped together into a single psychometric test score. It is our hope that this evolution in the orientation of the SLS will ultimately lead to further advances in our understanding of adult intellectual development.

ACKNOWLEDGMENTS

Research reported in this chapter was most recently supported by grants #AG-3544 and AG-4770 from the National Institute of Aging. Earlier phases of the study were supported by grants from the National Institute of Child Health and Human Development (HD-367 and HD-4476) and the National Institute of Aging

(HD-367). The enthusiastic cooperation of the staff and members of Group Health Cooperative of Puget Sound is gratefully acknowledged.

REFERENCES

Adam, J. (1978). Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin, 85*, 1309–1316.

Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. InD. J. Jackson & E. F. Borgatta (Eds.), *Factor analysis and measurement* (pp. 249–279). London: Sage.

Anastasi, A., & Drake, J. (1954). An empirical comparison of certain techniques for estimating the reliability of speeded tests. *Educational and Psychological Measurement, 14*, 529–540.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11*, 145–171.

Baltes, P. B., & Nesselroade, J. R. (1970). Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology, 2*, 244–256.

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., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Bayley, N., & Oden, M. (1955). The maintenance of intellectual ability in gifted adults. *Journal of Gerontology, 10*, 91–107.

Berg, C., Hertzog, C., & Hunt, E. (1982). Age differences in the speed of mental rotations. *Developmental Psychology, 18*, 95–107.

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectual des anormaux. *L'Anne Psychologique, 11*, 191.

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

Birren, J. E., Woods, A. M., & Williams, M. V. (1979). Speed of behavior as an indicator of age changes and the integrity of the nervous system. In F. Baumeister (Ed.), *Bayer Symposium IX: Brain function in old age* (pp. 10–44). Hamburg: Springer.

Bloxom, B. (1972). Alternative approaches to factorial invariance. *Psychometrika, 37*, 425–440.

Bock, D. R. (1975). *Multivariate statistical methods in behavior research*. New York: McGraw-Hill.

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.

Brooks, J., & Weintraub, M. (1976). A history of infant intelligence testing. In M. Lewis (Ed.), *Origins of intelligence* (pp. 19–58). New York: Plenum.

Carroll, J. B. (1976). Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 27–56). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carroll, J. B. (1980). *Individual difference relations in psychometric and experimental cognitive tasks*. Final Report No. 163, L. L. Thurstone Psychometric Laboratory, University of North Carolina.

Carroll, J. B., & Maxwell, S. E. (1979). Individual differences in cognitive abilities. *Annual Review of Psychology, 30*, 603–640.

Cerella, J., Poon, L. W., & Fozard, J. L. (1981). Mental rotation and age reconsidered. *Journal of Gerontology, 364*, 620–624.

Cooper, L. A., & Shepard, R. N. (1978). Transformations on representations of objects in space. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception*, Volume VIII: *Perceptual Coding* (pp. 105–146). New York: Academic Press.

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

Cronbach, L. J. (1970). *Essentials of psychological testing*. New York: Harper & Row.

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., & Owens, W. A., Jr. (1983). The Iowa State study of the adult development of intellectual abilities. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 20–39). New York: Guilford.

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

Egan, D. E. (1981). An analysis of spatial orientation test performance. *Intelligence, 5*, 85–100.

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

Embretson, W. S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179–197.

Galton, F. (1869). *Hereditary genius*. London: Macmillan.

Gaylord, S. A., & Marsh, G. R. (1975). Age differences in the speed of a spatial cognitive process. *Journal of Gerontology, 30*, 674–678.

Glenn, N. D. (1981). Age, birth cohort, and drinking: An illustration of the hazards of inferring effects from cohort data. *Journal of Gerontology, 36*, 362–369.

Gribbin, K., Schaie, K. W., & Parham, I. A. (1980). Complexity of life style and maintenance of intellectual abilities. *Journal of Social Issues, 36*, 47–61.

Hall, G. S. (1922). *Senescence, the last half of life*. New York: Appleton.

Hertzog, C. (1979). *A structural equations analysis of adult intellectual development*. Unpublished doctoral dissertation, University of Southern California.

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

Hertzog, C., & Schaie, K. W. (1982). *On the analysis of sequential data in life-span developmental research*. Paper presented at the 90th Annual Convention of the American Psychological Association, Washington, D.C.

Hertzog, C., & Schaie, K. W. (1984). *Age changes in intellectual structure: A structural equations analysis*. Unpublished manuscript.

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.

Hollingsworth, H. L. (1927). *Mental growth and decline: A survey of developmental psychology*. New York: Appleton.

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 1980s* (pp. 503–541). Washington: American Psychological Association.

Horst, P. (1954). A technique for the development of a differential prediction battery. *Psychological Monographs* (Whole No. 380).

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

Hunt, E. (1983). On the nature of intelligence. *Science, 219*, 141–146.

Hunt, E., Davidson, J., & Lansman, M. (1981). Individual differences in long-term memory access. *Memory and Cognition, 9*, 599–608.

Jarvik, L. F., Kallmann, F. J., & Falek, A. (1962). Intellectual changes in aged twins. *Journal of Gerontology, 17*, 289–294.

Jones, H. E., & Conrad, H. S. (1933). The growth and decline of intelligence: A study of homogenous group between the ages of ten and sixty. *Genetic Psychology Monographs, 13*, 223–298.

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 analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models*. Amsterdam: North Holland.

Jöreskog, K. G., & Sörbom, D. (1980). *Simultaneous analysis of longitudinal data from several cohorts*. Research Report 80-5. University of Uppsala, Dept. of Statistics.

Kail, R., Carter, P., & Pellegrino, J. (1979). The locus of sex differences in spatial ability. *Perception and Psychophysics, 26*, 182–186.

Kamin, L. J. (1974). *The science and politics of IQ*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kotovsky, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4*, 399–424.

Lansman, M., Donaldson, G., Hunt, E., & Yantis, S. (1982). Ability factors and cognitive processes. *Intelligence, 6*, 347–386.

Lohman, D. F., & Kyllonen, P. C. (1983). Individual differences in solution strategy on spatial tasks. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1). New York: Academic Press.

Mason, W. M., Mason, K. O., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review, 38*, 242–258.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins.

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

Mulaik, S. A. (1972). *Foundations of factor analysis*. New York: McGraw-Hill.

Nesselroade, J. R. (1970). Application of multivariate strategies to problems of measuring and structuring long-term change. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory* (pp. 194–207). New York: Academic Press.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression towards the mean and the study of change. *Psychological Bulletin, 88*, 622–737.

Owens, W. A., Jr. (1953). Age and mental abilities: A longitudinal study. *Genetic Psychology Monographs, 48*, 3–54.

Owens, W. A., Jr. (1959). Is age kinder to the initially more able? *Journal of Gerontology, 14*, 334–337.

Parham, I. A., Gribbin, K., Hertzog, C., & Schaie, K. W. (1975). *Health status change by age and implications for adult cognitive change*. Paper presented at the 10th International Congress of Gerontology, Jerusalem, Israel.

Pellegrino, J. W., & Glaser, R. (1980). Components of inductive reasoning. In R. E. Snow, P. A. Federocio, & W. E. Montague (Eds.), *Aptitude, learning, and instruction* (pp. 177–217). (Volume 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pellegrino, J. W., & Kail, R. (1982). Process analyses of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. (Volume 1). Cambridge: Cambridge University Press.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development, 15*, 1–12.

Pressey, S. L., Janney, J. E., & Kuhlen, R. G. (1939). *Life: A psychological survey*. New York: Hayer.

Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life span. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span psychological development: Research and theory* (pp. 468–484). New York: Academic Press.

Riegel, K. F. (1975). Adult life crises: Towards a dialectic theory of development. In N. Datan & L. Ginsberg (Eds.), *Life-span developmental psychology: Normative life crises* (pp. 99–128). New York: Academic Press.

Salthouse, T. A. (1980). Age and memory: Strategies for localizing the loss. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 47–65). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*, 92–107.

Schaie, K. W. (1973). Methodological problems in descriptive developmental research on adulthood and aging. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 253–280). New York: Academic Press.

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

Schaie, K. W. (1977/1978). Toward a stage theory of adult cognitive development. *Aging and Human Development, 8*, 129–138.

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, Jr. (Eds.), *Life-span development and behavior* (Vol. 2) (pp. 67–115). New York: Academic Press.

Schaie, K. W. (1980). Intelligence and problem solving. In J. E. Birren & R. B. Sloane (Eds.), *Handbook of mental health and aging* (pp. 262–284). Englewood Cliffs, NJ: Prentice-Hall.

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. (1984). Historical time and cohort effects. In K. A. McCloskey & H. W. Reese (Eds.), *Life-span developmental psychology: Historical and generational effects* (pp. 1–15). New York: Academic Press.

Schaie, K. W., & Baltes, P. B. (1975). On sequential strategies in developmental research: Description or explanation. *Human Development, 18*, 384–390.

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 studies 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 edition). New York: Van Nostrand Reinhold.

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

Schaie, K. W., Labouvie, G. V., & Buech, B. U. (1973). Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. *Developmental Psychology, 9*, 151–166.

Schaie, K. W., & Labouvie-Vief, G. (1974). Generational versus ontogenetic components of change in adult cognitive behavior: A fourteen-year cross-sequential study. *Developmental Psychology, 10*, 305–320.

Schaie, K. W., & Parham, I. A. (1977). Cohort-sequential analyses of adult intellectual development. *Developmental Psychology, 13*, 649–653.
Schaie, K. W., Rosenthal, F., & Perlman, R. M. (1953). Differential deterioration of factorially "pure" mental abilities. *Journal of Gerontology, 8*, 191–196.
Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin, 70*, 672–680.
Siegler, I. C. (1983). Psychological aspects of the Duke longitudinal studies. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 136–190). New York: Guilford.
Sinott, J. D. (1981). The theory of relativity: A metatheory for development? *Human Development, 24*, 293–311.
Sorbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.
Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Sternberg, R. J. (1980). Sketch of a componential subtheory of human intelligence. *The Behavioral and Brain Sciences, 3*, 573–614.
Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38*, 878–891.
Stone, V. (1980). *Structural modeling of the relations among environmental variables, health status, and intelligence in adulthood*. Unpublished doctoral dissertation, University of Southern California.
Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton.
Thurstone, L. L. (1938). *The primary mental abilities*. Chicago: University of Chicago Press.
Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.
Thurstone, L. L., & Thurstone, T. G. (1949). *Examiner manual for the primary mental abilities test (Form 11-17)*. Chicago: Science Research Associates.
Waugh, N. C., & Barr, R. A. (1980). Memory and mental tempo. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 251–260). Hillsdale, NJ: Lawrence Erlbaum Associates.
Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
Wilkie, F. L., & Eisdorfer, C. (1971). Intelligence and blood pressure in the aged. *Science, 172*, 959–962.
Willis, S. L. (1985). Towards an educational psychology of the adult learner. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*. New York: Van Nostrand Reinhold.
Willis, S. L., & Schaie, K. W. (1981). Maintenance and decline of adult mental abilities: II. Susceptibility to experimental manipulation. In F. Grote & R. Feringer (Eds.), *Adult learning and development* (pp. 40–57). Bellingham, WA: Western Washington University.
Willis, S. L., & Schaie, K. W. (1983). *Enhancing intellectual performance in well functioning elderly*. Paper presented at the annual meeting of the Gerontological Society of America, San Francisco.
Yerkes, R. M. (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences, 15*, 1–890.