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

7

Longitudinal Methods

K. WARNER SCHAIE
CHRISTOPHER HERTZOG

INTRODUCTION

Investigators studying problems in the developmental sciences, particularly developmental psychologists, have long felt that the explication of lawful relations in developmental processes of necessity requires observing the same organisms over that period of time during which developmental phenomena of interest are likely to occur. However because most developmental phenomena of interest in humans occur relatively slowly (with the exception of early infancy and the period prior to death in old age), it is not generally practical for developmental research designs to follow subjects over the entire developmental period. Thus a variety of developmental research designs have been devised to finesse the problem by assuming that some estimate of developmental change from an experimental or quasi-experimental design may be substituted for the long-term observation. Many investigators have tried to model developmental phenomena with more economical designs, whether by experimental induction of change, by retrospective analysis, or by comparisons of individuals of differing developmental levels at one point in time. These more economical developmental designs all have merit depending upon whether the assumptions necessary to apply the design are accurate and whether the design is best suited for the specific questions to be asked. It is often the case that the status of these designs as quasi-experiments is ill defined and that the enabling assumptions have not been explicitly formulated.

It is the purpose of this chapter to discuss three topics involved in the correct understanding and application of research design method to the study of developmental processes. The first topic places the various developmental methods within the context of the broad group of quasi-experimental designs (Campbell & Stanley, 1967) useful in developmental research. The second topic concerns explication of the interdependence of developmental theory and selection of the best suited research designs, such that data collection and analysis are properly derived from theory and thus capable of testing theory-derived hypotheses. The third topic involves the description of statistical models and estimating techniques currently available for the modeling of developmental processes. Here special emphasis will be given to the recently developed methods permitting the application of factor analysis and linear structural equation systems to developmental data.

As will become clear upon reading this chapter, these three topics are interrelated in many ways and a number of different approaches could be chosen for their presentation. We begin with a discussion of the relation between developmental theory and longitudinal methods, including a set of definitions and principles, a discussion of the advantages and disadvantages of the longitudinal methods, and some guidelines for the suitability of alternate theoretical models in attaining parsimonious design choices. We next explicate the types of quasi-experimental designs common in developmental psychology. In this context we consider the methodological problems intrinsic to the common developmental designs, present expanded sequential strategies to reduce limitations of the more traditional developmental designs, consider the remaining confounds and their impact, and assess the applicability of proposals for the simultaneous estimation of confounded parameters coming from sociological research. The final section on statistical techniques for lon-

Preparation of this chapter was greatly facilitated by research support of the first author from the National Institute of Aging (AG-480) and a postdoctoral fellowship support for the second author, also from the National Institute of Aging (AG-5150).

gitudinal analysis briefly highlights methods for the test of hypotheses concerning means and covariance structures. Space limitations prevent us from elaborating on these methods in any detail, but the interested reader is directed to the relevant reference materials.

DEVELOPMENTAL THEORY AND LONGITUDINAL METHODS

Definition and Principles

In general, the goal of longitudinal methods in developmental psychology is to obtain valid measures of developmental change for descriptive and explanatory purposes. This section will attempt to link the application of longitudinal methods with underlying developmental theory by highlighting current views of development as a time-dependent process, indicating alternate models of development, and offering some definitions of what is or is not included in the term "longitudinal methods."

Development as a Time Dependent Process. Although development consists of intraindividual change over time, not all such change should, in principle, be developmental; indeed, there has been some controversy as to the attributes which would enable one to posit behavioral change as truly *developmental* (cf. Baer, 1970; Baltes & Nesselroade, 1979; Baltes & Willis, 1977; Reese & Overton, 1970). *Organismic* conceptions of development demand that developmental change have the attributes of universality, fixed sequentiality, structural and qualitative transformations and orientation toward an end state (see McCall, 1977, or Wohlwill, 1973 as representatives of this position). Such *strong* conception of development is contrasted by the operant position which might be characterized as a *weak* model. Here in its extreme form any reliably observed form of behavior change might be viewed as development (cf. Baer, 1970; Bandura, 1971). Generally, however, most developmental researchers would agree with Baltes and Nesselroade (1979) that there are some minimal criteria needed to label change as being developmental. They suggest "one needs a theory-based or empirically derived behavior change process on the descriptive level . . . [and] . . . the use of historical time-ordered paradigms of influences for the explanation of developmental change" (p. 15).

We may start then with Kessen's (1960) formulation that "a characteristic is said to be developmental if it can be related to age in an orderly way" (p. 36), or $B = f(A)$. Kessen also specified that a response in a developmental model should be seen generally as a function of age as well as a special population and an environment. In earlier work (Schaie, 1965, 1973) we have explicated this expanded notion by a model where $B = f[A,C,T]$. It should be noticed that the three items involved are strictly *descriptive* parameters. Their definition implies that *age* $[A]$ will refer to the number of time units elapsed between the birth (entrance into the environment) of the organism and the point in time at which the dependent variable is measured. The special population in this expanded model is generally defined as *cohort* $[C]$, which implies the total population of individuals entering the specified environment at the same point or interval in time.[1] Environment is for descriptive purposes more precisely defined by the term *time of measurement* $[T]$ (equivalent to the term *period* in the sociological literature)—that is, the point in time at which the response of interest is actually recorded.

The critical problem in working with the model $B = f[A,C,T]$ is that although the three effects are conceptually independent, the three variables are in fact linearly dependent. If we define C by birth year, then $T = C + A$; if year of birth and age are known, then the current year may be specified with complete certainty. As we shall see, the research designs to be discussed in this chapter aid in the estimation of these three effects, but they do not eliminate the indeterminacy due to the linear dependence except by making assumptions about the presence or absence of some of the effects in question. Thus application of the methods described here may, in principle, disentangle the effects of A, C, and T (although the validity of the approach depends directly on the veracity of the assumptions which must be made to enable estimation of the various effects).

It cannot be emphasized enough that a descriptive model based on these methods can do nothing more than identify those effects which are *correlated* with A, C, or T, independent of other parameters. Given an organismic conception of development, we cannot conclude that effects associated with A, and not C or T, represent developmental change in the strong sense referred to above unless there are theoretically sound justifications and additional evidence for concluding that the age-correlated change is truly representing ontogenetic development. From a life-span perspective the justification for inferring ontogenetic change from significant A effects might be generally stronger in child development than in adult development, although this position would be vigorously debated by child psychologists and gerontologists alike. In any case one cannot conclude that A effects imply inevitably the existence of ontogenetic change. For example a study of attitudes and life-satisfaction after age 65 would probably be able to isolate age-correlated change, common to a range of cohorts and times of measurement. Yet such change might not necessarily be a function of ontogenetic development since it could also be specific to retirement and related to A only by virtue of the fixed retirement age prevalent in the society over the period of measurement. As such the changes might well be due to an age-correlated but event- or sequence-relevant phenomenon (Baltes & Willis, 1977).

It follows then that developmental analysis only begins with the parametric description of age-correlated change, and understanding of development also requires an attempt to provide explanations of behavioral development from a process-oriented perspective. A process-oriented explanation of development requires specification of antecedent-consequent relationships at a causal

[1] It should be noted that cohort can also be defined by entry into a common environment by individuals of different ages, e.g., a college class or the initial work force of a new factory. Restrictions regarding the nature of the population and latitude in defining the boundaries of a given cohort will depend upon the special assumptions appropriate to the problem being investigated (see also Rosow, 1978).

level. The specification and testing of alternative process-oriented models of behavioral development is obviously a principal goal of developmental psychology and yet is its most difficult task, since developmental phenomena are generally produced by complex interactions among a set of underlying processes which may produce a range of developmental phenotypes depending upon the environmental milieu in which their genotypes operate. Thus validation of process-oriented models of development must inevitably require the analysis of a multivariate system of variables measuring the behavioral domain of interest and the putative processes determining developmental change in this domain (Baltes & Nesselroade, 1973; Nesselroade, 1977).

It is rather implausible to assume that behavioral development could be encapsulated in any univariate causal model. Much of our discussion of longitudinal methods must then focus on the design and analysis of behavioral investigations from a multivariate perspective as it affects measurement of relevant constructs and specification of causal relations among these constructs.

In order to begin to provide an accurate descriptive and explanatory account of developmental phenomena, we must begin by specifying a set of competing models which account for these phenomena. These models must obviously be directed toward an explanation of change within individuals (intraindividual change), as emphasized by most developmental theorists (e.g., Baltes & Nesselroade, 1979; Wohlwill, 1973). The models must also allow for differences among individuals at any given point in time (interindividual differences) and interindividual differences in the course of intraindividual change (see Baltes & Nesselroade, 1979; Baltes & Willis, 1977). The power of such models will be directly related to their ability to account for systematic age-related intraindividual change *and* interindividual differences in intraindividual change on the basis of a set of causal processes varying between individuals. When all is said and done, we are concerned with deriving estimates for populations and their subtypes, as well as definitions of the range within which intraindividual variability about such parameters may be found (for further discussion of these issues see Baltes & Nesselroade, 1979; Buss, 1974).

Models of Development. The selection of a longitudinal design for a developmental study depends critically upon the developmental model believed applicable to the phenomenon (Schaie, 1973). Thus it is important to consider the (often implicit) developmental models used by behavioral scientists and their salient characteristics vis-à-vis design selection.

A distinction must first be drawn between developmental models implying either quantitative or qualitative change. Quantitative change implies continuous incremental or decremental change in some measures of behavior, given the assumption that the underlying processes determining the behavior remain fixed or static. Quantitative models of development need not be linear (and might even be recursive); the assumption of linearity and unidirectionality is usually made only as a simplifying convenience for data analysis. The critical factor in a quantitative model is that development is represented only in changes in the performance level on some behavioral scale. By contrast qualitative change is often considered to require discrete shifts from one stage to another (e.g., the Piagetian cognitive stages); qualitative developmental change implies that behavior at a later stage of development cannot be accounted for as a simple function of behavior at an earlier state of development—rather some structural metamorphosis among determining processes is hypothesized to have produced a new function relating process to behavior (see Wohlwill, 1973, for a useful account of these concepts and of criteria for inferring qualitative change).

Developmental models are not necessarily exclusively quantitative or qualitative in nature since in qualitative models continuous changes are said to occur within a given stage until a threshold level is reached, at which point the transformation to the next stage occurs. As pointed out by Baltes and Nesselroade (1970, 1973), the distinction between qualitative and quantitative developmental change is critical in determining the adequacy of longitudinal methods in a given application. It makes little sense to make purely quantitative comparisons of behavior on a given performance measure at two different levels of development if qualitative change has occurred because the likelihood is high that qualitative change will impose a qualitative difference in the relationship of the behavioral scale to the underlying constructs the scale supposedly measures.

If we expect qualitative differences in antecedent-consequent relationships due to development, our analysis must be designed concomitantly to characterize the nature of the qualitative change and to estimate the extent of quantitative change over the age range where the process-behavior relationships may be assumed static and thus characterized as a continuous quantitative function. Whether our model is quantitative or qualitative, it is obvious that the ages of development studied in our model will have implications as to the kind of data needed for full description of the developmental process. With respect to development during childhood, it is usually true that, where quantitative models are sufficient to account for a developmental phenomenon, an *incremental* model will fit most variables. The incremental model implies monotonic increases in performance level on a behavioral scale, although the function describing the behavioral increment may not be linear but rather a slowly decelerating growth rate and young adult asymptote as specified by a Gompertzian growth curve. Whether a continuous or stage model is assumed to apply to childhood development, interindividual differences about the normative developmental rates and the temporal latitude of stages are generally assumed to be narrow.

Studies of adult development have, for the most part, implicitly assumed an *irreversible decrement* model. This model, common to analyses in areas such as intelligence, creativity, and achievement, assumes that maximal level of function is reached in young adulthood followed by a linearly accelerating and irreversible decline. The irreversible decrement model implicitly assumes that adult development may be characterized by a purely quantitative model of decline secondary to the aging process; it generally emphasizes the process of biological decrement with aging (Baltes & Willis, 1977; Schaie, 1973). Another implicit specification of the irreversible decrement model

is that age changes occur as a function of maturational (ontogenetic) events which are affected but slightly by environmental variation.

Research in adult psychological development has been dominated by the irreversible decrement model, to the exclusion of other models of potential explanatory power (Baltes & Willis, 1977; Barton, Plemons, Willis, & Baltes, 1975; Labouvie-Vief, 1977; Labouvie-Vief & Chandler, 1978). It has generally been the case that age-correlated effects are assumed to be evidence of irreversible decrement, even in the absence of corroborative evidence that processes other than biological decrement due to aging cannot account for the age-correlated effects. The irreversible decrement model is a valid model for some psychological processes, most notably those functions which are directly related to the biological integrity of the central nervous system; however, we must echo the concerns of Baltes and Willis (1977) and others that research in adult psychological development should not axiomatically postulate the irreversible decrement model as a theoretical basis for all research questions of interest.

Given the position that alternative models for adult psychological development other than the irreversible decrement model should be developed and tested, what alternative models might be specified? Two simple alternatives (Schaie, 1973) are the *adult stability* and the *decrement with compensation* models. The adult stability model postulates that once an adult asymptote is reached, behavior remains stable throughout the remaining life span. However cyclical changes might still occur about an optimal level as the result of both external and internal events (cf. Goulet, Hay, & Barclay, 1974; Schaie, 1973). The stability model has been assumed to hold in the study of personality traits, and it may also fit components of cognitive development such as crystallized intelligence.

The decrement with compensation model, increasingly popular as a result of gerontological intervention studies, expects decline past maturity but argues that environmental intervention may compensate for maturationally programmed deficits. This model might fit concepts such as fluid intelligence or measures of performance where decline is to be expected due to correlated decremental biological events but where environmental input might have significant moderating effects. An excellent example in the area of physiology is the research by DeVries (1974) which suggests that programmed exercise for the elderly may ameliorate muscular and cardiovascular decline previously assumed to be irreversible.

Another important model for adult development is the *sequence-relevant* model (Baltes & Willis, 1977), which posits that age-correlated change is specific to a programmed sequence of processes which are correlated with age but not isomorphic with ontogenetic change—that is, a sequence of processes which are not inevitably associated with the aging process. One of the implications of the sequence-relevant model is that developmental psychologists must differentiate normative from nonnormative age change (cf. Baltes & Nesselroade, 1979), which can be an extremely difficult theoretical problem. Nevertheless the distinction of sequence-relevant change from age-relevant change is an important one for it is quite possible that effects which have been assumed to be a function of ontogenetic decline might be better described by an adult stability model coupled with a probabilistic occurrence of sequence-relevant change. An example would be the terminal decline or terminal drop hypothesis of Riegel and others (e.g., Riegel & Riegel, 1972; Siegler, 1975) which posits stability of adult intelligence until impending death when physiological pathology compromises the functional efficiency of the nervous system.

Finally we must recognize that no one developmental model need be valid for all individuals. One of the characteristics of human development is the ever-increasing range of individual differences; thus developmental psychologists need to specify models which take into account the wide variety of multidimensional and multidirectional possibilities for patterns of developmental change. It is unlikely that a life-span–oriented approach can, for any variable system, sufficiently account for development with a single developmental model. Alternate patterns of change are highly likely for the stages of late maturation, adulthood, and senescence.

Longitudinal Methods. The term *longitudinal methods* has been used in a variety of ways. Hindley (1972), for example, claims that "there is no hard and fast definition of what constitutes a longitudinal study" (p. 23), although Baltes and Nesselroade (1979) contend that one requirement of a longitudinal inquiry must be that "the entity under investigation is observed repeatedly and evolves over time" (p. 4). For our purposes, however, we would like to include within the general category of longitudinal methods at least that variant of a longitudinal study which does not involve repeated measurement of the same individual—namely, sampling procedures in which a cohort is observed repeatedly by means of successive random samples from the parent population.

Longitudinal methods traditionally involve age-based parametric models in which chronological age is the predictor variable of central importance. However it would be quite feasible to include designs which might address hypotheses where there is a directional time sequence, even one which is uncorrelated with chronological age. Indeed such designs might well be required in order to address questions as to the cross-cultural congruence of universal developmental stages. Thus our discussion of longitudinal methods will include designs which are not technically "longitudinal" by the stricter criteria of others; we consider designs which are longitudinal only in the sense that they model age-correlated effects by sampling across a sequence of points.

In the past, longitudinal methods have been utilized primarily for the purpose of describing developmental phenomena, and much of our discussion of longitudinal method focuses on descriptive applications. However these methods can be readily applied to explanation and intervention by specifying prediction systems of process variables or introducing design extensions which can handle treatment effects (e.g., Labouvie, 1974, 1978).

Advantages of Longitudinal Methods

The primary advantage of the longitudinal methods, of course, is the fact that they emphasize intraindividual change (IAC) while cross-sectional approaches can make

statements only about interindividual variability (IEV). Even in the case where independent samples are studied over time from a given cohort, the emphasis is then on change within the populations examined rather than upon differences between samples possibly coming from noncomparable populations. Obviously most longitudinal approaches permit analyses of IEV in addition to IAC.

Following Baltes and Nesselroade (1979) we can identify five distinct rationales for longitudinal studies of behavioral development. Of these, three involve developmental descriptions while the other two are explanatory in nature. As indicated previously, the first rationale is concerned with the *direct identification of IAC.* Such change can be quantitative and continuous, or it can involve transformation of one behavior to another or changes in the patterns of observed variables as they measure theoretical constructs. Observations based on a single occasion are simply not appropriate for this purpose. To be explicit, if cross-sectional data are to be used to estimate IAC, the assumptions to be met would include that (1) subjects must be matched across age levels, (2) different-aged subjects must come from the same parent population at birth, and (3) different-aged subjects must have experienced identical life histories. Such assumptions cannot be met in human studies.

The second rationale concerns the direct *identification of IEV in IAC*—that is, we are here interested in the degree of variability displaced by different individuals in their behavioral course over time. Examination of similarities and differences in developmental patterns requires the availability of measures of longitudinal change *within* individuals. Unless such data are available, it is not possible to answer the question of whether or not group parameters are characteristic of the development of any individual. Of course the valuable hypothesis-generating source of single-subject research depends upon longitudinal analyses (cf. Shontz, 1976).

Third, longitudinal data permit the *analysis of interrelations among IAC* within a multivariate behavioral domain of variables. Only when several individual behaviors have been followed over time is it then possible to discover constancy and change of the entire organism, particularly where a wholistic or structural approach is taken to human development (e.g., Riegel & Rosenwald, 1975). Longitudinal studies alone, by means of multiple observations over time, permit the discovery of structural relations among changes in behavior. Such approaches are obviously essential for the meaningful identification of systems and progressive differentiation processes as essential concepts in the understanding of human growth and development (cf. Lund, 1978; Urban, 1978).

The fourth rationale for longitudinal studies involves the *analysis of determinants of IAC.* Here we are concerned with the identification of time-ordered antecedents and consequents as necessary, albeit not sufficient, conditions for causal inference. Longitudinal data alone can provide the necessary data when the causal process involves discontinuity (e.g., sleeper effects), is multidirectional, or contains a multivariate pattern of influences (cf. Baltes, Reese, & Nesselroade, 1977; Heise, 1975).

Fifth and finally, longitudinal studies permit the *analysis of IEV in the determinants of IAC.* What is at issue here is the fact that many individuals can show similar patterns in intraindividual change which may be determined by different change processes. This may be the case for persons at different levels in the range of talent or other personality attributes. But interindividual differences in patterns of change may also be due to the operation of alternate combinations of causal sequences.

Importance of Strong Developmental Theory

The inherent confounds implicit in the longitudinal methods require strong, clearly specified developmental theories in order to generate meaningful hypotheses and, in fact, to permit design economies which make certain longitudinal inquiries logistically feasible. Some of the major objections of recent papers critical of Schaie's sequential methods (Adam, 1978; Buss, 1979–1980) are related primarily to the conceptual difficulties an investigator encounters when attempting to apply sequential strategies using a totally atheoretical approach. While we will consider these strategies and their limitations in detail, it should be mentioned here that it is certainly unprofitable for developmental psychologists (or any other scientists) to pursue descriptive paradigms in a theoretical vacuum; such an approach is not merely weak science but highly prone to the "discovery" and perpetuation of misleading inferences.

The critical problem (one which, quite frankly, remains to be satisfactorily solved) is how plausible rival models of development may be tested when confounds inherent in longitudinal methods appear to preclude designs which make the alternative model's assumptions directly falsifiable. The following methods do not resolve this problem; however, we would argue that they provide a method of matching a developmental model to research design in a way which produces valid estimates of hypothesized effects, given that the model is valid and, further, that their assumptions are less restrictive (and more likely to hold true) than the assumptions enabling the simpler, more traditional designs. The developmental psychologist should be acutely sensitive to the fact that the parameter estimates from a chosen design are no more valid than the model assumptions which permit their estimation. There is simply no substitute for an explicitly defined developmental model; even a misspecified model, if explicitly formulated, at least provides a basis for understanding the potential consequences of the misspecified sources of effects.

LONGITUDINAL METHODS AS QUASI-EXPERIMENTAL RESEARCH DESIGN

Multiple Observations with or without Experimental Treatments: Methodological Problems

Internal and External Validity. Because age is a subject attribute which cannot be experimentally assigned (at least not without a time machine), longitudinal studies cannot conform to the rules for true experiments and hence are subject to all the problems inherent in what Campbell and Stanley (1967) term *quasi-experiments.*

These problems may be categorized as either threats to the *internal validity* or the *external validity* of a given quasi-experiment. Internal validity is upheld if the factors analyzed in a given design are truly measures of the hypothesized construct and are not confounded by other factors not explicitly included in the design. External validity defines the limits of valid generalization from the findings of a given study.

Campbell and Stanley (1967) enumerated eight different threats to the internal validity of a pretest-posttest design: effects of history, maturation, testing, instrumentation, statistical regression, mortality, selection, and the selection-maturation interaction. For the developmental psychologist, history and maturation have special meaning above and beyond the internal validity threat posed for a pretest-posttest design. Maturation is quite obviously not a threat in developmental studies but rather the specific variable of interest. The fact that maturation is the primary effect of interest to developmental psychologists does not imply that the measurement of maturational effects is inevitably straightforward; given a specific developmental model, it may be crucial to not merely test the null hypothesis of no maturational effects but rather some explicit alternative hypothesis specifying the direction and magnitude of the expected maturational effect.

Historical effects, on the other hand, are the primary source of internal validity problems for the developmental psychologist. History is directly tied to both cohort and time-of-measurement effects. A cohort, as we have defined it, is a group of individuals born in the same historical period who therefore share the same environmental circumstances at the same point in their maturational sequence. Time-of-measurement effects represent the events which affect all members of the population living at a given period of history. In both cases historical events may modify the range of person-environment interactions and limit the external validity of any internally valid estimate of maturational change. However, as we shall see shortly, historical effects, operating as either cohort or time-of-measurement effects, may threaten the internal validity of designs attempting to measure maturation per se.

Since the traditional longitudinal design is a special case of the pretest-posttest design in that it repeatedly measures the same individuals over time, the other six internal validity threats listed by Campbell and Stanley are important threats to the validity of longitudinal designs as well. The validity threats are discussed in greater detail later, but we now supply their definitions: *Testing* refers to the effects of the measurement process itself, which may be confused with maturational effects. There are two major effects of testing per se—practice and reactivity. The act of testing itself provides practice on the test, which should in general lead to improvements in performance with each new retest. *Reactivity* refers to the possible effects of being tested on subsequent behavior because the subjects react to being tested by behaving differently than had they not been tested. Such effects could also be confused with maturation. *Instrumentation* refers to any differences in the measurement techniques which covary with the measurement occasions. *Statistical regression* refers to the tendency for variables containing measurement error to regress toward their mean from one occasion to the next. *Mortality* refers to the attrition of subjects from a sample between measurement occasions; it is termed *experimental mortality* so as to include attrition due to biological mortality, morbidity, and other psychological and sociocultural factors. *Selection* refers to the process of obtaining a sample from the population, and the *selection-by-maturation interaction* refers to the possibility that variation in the method of sample selection may produce variation in the maturational effects to be estimated (see also Cook & Campbell, 1975).

In addition to the threats to the internal validity of quasi-experiments, Campbell and Stanley (1967) call attention also to a number of limitations (threats to external validity) with respect to how widely findings from such studies can be generalized. These limitations are concerned with questions regarding the *experimental units,* the extent to which longitudinal data collected on one sample can permit inference to other populations; *experimental settings,* the extent to which findings have cross-situational validity (cf. Scheidt & Schaie, 1978); *treatment variables,* limitations imposed by specific settings of measurement-implicit reinforcement schedules (cf. Birkhill & Schaie, 1975; Schaie & Goulet, 1977); and *measurement variables,* the extent to which task characteristics are appropriate at different developmental stages in a longitudinal study (cf. Schaie, 1977–1978; Sinnott, 1975).

Longitudinal methods as defined in this paper are generally designed to estimate the expanded function $B = f[A,C,T]$ in an economical manner. As discussed previously by many (e.g., Kessen, 1960; Schaie, 1965), developmental psychologists cannot afford to wait a lifetime to produce answers to the research questions that interest them. The problem is particularly acute for life-span studies of human development, where explication of A, C, and T parameters over a wide range of C and T values would require the impossible: that the experimenter outlive his or her subjects by at least one (or more) lifetimes! Hence one needs designs which compromise the conflicting goals of maximal external validity over possible variables in A, C, and T and minimal investment of time in data collection.

The most economical design in terms of time investment is the simple *cross-sectional* design, which samples a range of individuals of varying chronological ages at a single point in time. In the simple cross-sectional design, too much may have been given up in the name of economy since the effect of A is completely confounded with C (cf. Schaie, 1965). Thus the estimates of C obtained from the cross-sectional design are internally invalid unless the strong assumption of no effects for A may be made (Baltes, 1968; Kuhlen, 1963; Schaie, 1965, 1973, 1977).

The cross-sectional design, when applied to the study of development, represents an attempt to estimate IAC functions from IEV data which, taken in isolation, may result in incorrect inferences about developmental functions. Thus the simple cross-sectional design is not the design of choice for developmental research. Given that the focus of this chapter is upon longitudinal methods, we shall have little else to say about the simple cross-sectional design. Our attention now turns to the discussion of the economy and validity problems of other developmental designs.

Traditional, Single-Cohort Longitudinal Designs. The classic longitudinal design was developed for the purpose of explicitly estimating development as IAC—emphasizing that the most valid estimates of development measure change over time in the same individuals. Explicitly this design represents a time series with an initial pretest, a subsequent intervention (traditionally the maturational events occurring over time), and a posttest, all on the same individual organisms. If the longitudinal study is continued over more than a single time interval, there is simply a further succession of alternating treatments (read maturational events) and further posttests. Traditionally the longitudinal design was only applied to one group of individuals of relatively homogenous chronological age at first testing and, therefore, to a single birth cohort.

As pointed out repeatedly (Schaie, 1965, 1972a, 1973, 1977) the single-cohort longitudinal design is highly susceptible to validity threats and should be avoided unless (1) experimental isolation can be achieved or (2) it can be shown that the dependent variable is not influenced by external environmental events. Barring such strong assumptions (which will rarely hold for studies of human development), several of the threats to internal validity enumerated above are likely to provide alternative explanations for the observed behavioral change (or lack thereof) which are as plausible as sources of the effect as is maturation itself. First, T and A are completely confounded, and thus any period effects related to the dependent variable will render estimates of A internally invalid. These period effects may either mimic or suppress maturational changes occurring over the particular age span measured depending upon whether A and T covary positively or negatively. Second, the single-cohort longitudinal design does not directly control for other internal validity threats which plague test-retest designs—namely, testing, instrumentation, experimental mortality, and statistical regression. The careful researcher can eliminate the confound of instrumentation by taking steps to assure that the measurement procedures are as consistent as possible, and statistical-regression effects can be minimized by including at least two retest occasions (Baltes & Nesselroade, 1979); however, there is simply no way for the single-cohort longitudinal design to circumvent the confounds of testing and experimental mortality effectively. Repeated testing must inevitably introduce the possibility of practice effects or reactivity to the testing situation, and the requirement of multiple test occasions virtually insures an attrition of some subjects who participated at the initial testing but who are unavailable, for whatever reason, for subsequent retesting. To the extent that experimental mortality simply produces a positively biased sample, then the problem is one of external validity—i.e., the overall level of performance of the attrited sample is higher than that of the population. However if the attrition effect interacts with maturation such that returning subjects have different developmental functions than attrited subjects, then the developmental function estimated by the single-cohort longitudinal design is internally invalid.

The single-cohort sampling of the traditional longitudinal design also limits the external validity of the design. Given that cohorts may differ in person-environment interactions due to historical effects (e.g., the children of the great depression; cf. Elder, 1974), descriptions of maturational change derived from a single cohort may well be unreplicable (i.e., externally invalid) for other cohorts.

Considering the problems inherent in the single cohort longitudinal design, one of us has explicitly wondered whether this design is ever a completely valid design for developmental research (Schaie, 1972a), a position which has evoked a spirited defense of traditional longitudinal methods (McCall, 1977). In fairness we should emphasize that the single-cohort longitudinal design, while deficient as a general method for studying developmental phenomena, may prove useful in particular applications, such as defining typologies of developmental patterns for a specifically targeted, single cohort population. Moreover the single cohort design may provide preliminary evidence regarding developmental functions which will later be replicated for additional cohorts and measurement occasions. Exclusive use of the single cohort longitudinal design for discovering normative laws of development is ill-advised unless the (probably unrealistic) assumptions of no period effects and no cohort-by-maturation interactions can be theoretically justified, a priori, for a specific population and behavior.

Sequential Strategies

Definitions. In order to reduce the limitations inherent in the single-cohort longitudinal design, several alternative *sequential* strategies have been suggested (Baltes, 1968; Schaie, 1965, 1970, 1977). The term *sequential* derives from the fact that the sampling frame for these designs requires a sequence of samples taken across multiple measurement occasions. In order to explicate the various possible sequential designs, we must first differentiate the *sampling design* from the *analysis design.* The two concepts are heavily interrelated; sampling design refers to the particular cells of a cohort-by-age (time) matrix to be sampled in a developmental study, while analysis design refers to the ways in which the cells which have been sampled may be organized to analyze for the effects of A, C, and T. Figure 7-1 provides a prototypical cohort-by-age matrix, which may be used to illustrate the various sequential designs. Note that, given the inherent confounding of A, C, and T discussed earlier, the cohort-by-age matrix in Figure 7-1 represents all three parameters: A and C as rows and columns of the matrix, and T as a parameter contained within the cells of the matrix.

We may distinguish two types of sampling designs: those which use repeated measurements on the same individuals to fill the cells of the matrix and those which use independent samples of individuals to fill the cells of the matrix; either sampling method could be used to provide the matrix given in Figure 7-1. Restricting our discussion (for the present) to the sampling design, we may use Baltes's (1968; Baltes, Reese, & Nesselroade, 1977) terminology to define these two designs as *longitudinal* and *cross-sectional sequences.* As shown in the diagonals in Figure 7-1, a cross-sectional sequence involves the replication of a cross-sectional study in that the same age range of interest is assessed for at least two different time periods. As a consequence the estimate for each age level

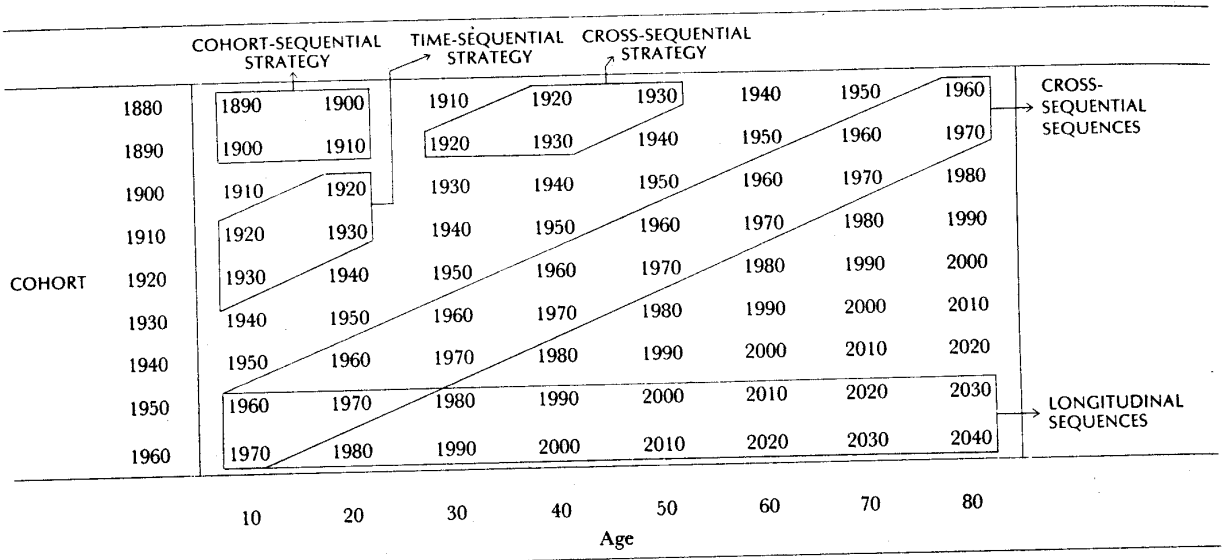

Figure 7-1 Schematic showing cross-sectional and longitudinal sequences and the modes of analysis deduced from the general developmental model. (Note: Entries represent times of measurement [period].)

is obtained for multiple cohorts. Each estimate, however, is obtained from an ideally random sample of its age-cohort and measured only once. By contrast the longitudinal sequence (bottom rows in Figure 7-1) represents the repeated measurement of at least two cohorts over the age range of interest. Here again estimates from each cohort are obtained at two different points in time. A critical distinction is the fact that only the longitudinal sequence provides data which permit evaluation of IAC and IEV in IAC.[2]

Schaie's General Developmental Model. It was pointed out earlier (Schaie, 1965) that data matrices like Figure 7-1 contained information permitting a variety of alternate strategies of analyses.[3] To be specific, each row can be treated as a longitudinal study, each diagonal as a cross-sectional study, and each column would represent a time-lag study (i.e., comparison of behavior at a specific age for successive cohorts). The sequential sampling designs cannot disentangle all components of the $B = f[A,C,T]$ function, given the linear dependence among the three factors. However, Schaie (1965) suggested that, given the $B = f[A,C,T]$ model, three distinct analysis designs exist which are created by considering the separate effects of two of the three components while assuming the constancy or irrelevance of the third. Consequently we suggested (as exemplified in Figure 7-1) that the *cohort-sequential* strategy would permit separation of age change from cohort differences, under the assumption of no time effects; the *time-sequential* strategy permits the separation of age differences from period differences, assuming no cohort effects; and the *cross-sequential* strategy permits the separation of age-cohort differences from period differences, assuming no age effects. The time-sequential strategy is not a truly longitudinal approach (i.e., one cannot do a time-sequential analysis on repeated measures data, since a given individual cannot be the same age at two different points in time—see Figure 7-1) and will not therefore be considered further except to note that it has merit for the estimation of reliable age differences, for social policy purposes, and for dependent variables where cohort effects are likely to be minimal.

Longitudinal Sequences. When data are collected in the form of longitudinal sequences (which should rightly be emphasized in developmental studies of IAC), it is possible to apply both the cohort sequential [*CS*] and cross-sequential [*XS*] designs. There is now general agreement that the *CS* design is of greatest interest to developmental psychologists because it explicitly differentiates IAC within cohorts from IEV between cohorts (cf. Baltes & Nesselroade, 1979; Schaie & Baltes, 1975). Not only does the *CS* design disentangle the effects of *A* and *C*, completely confounded in simple cross-sectional designs, but it also permits a check of the consistency of age functions across successive cohorts, thus proving to be of greater external validity than the single cohort longitudinal design.

Again the critical assumption in the *CS* design is that there are no time-of-measurement effects present in the data; this assumption is most parsimonious for developmental psychologists for whom age and cohort are likely to be of primary interest. Nevertheless there may be period-specific effects present in the data, either because of "true" period effects or because of confounding of occasion-related internal validity threats such as differences in instrumentation between occasions. We may therefore

[2] It would, of course, be possible to construct row entries in Figure 7-1 by means of successive independent samples from the cohorts under observation. Such data would use age correlated IEV to provide estimates on averaged IAC but would no longer be "truly" longitudinal—i.e., repeated observations on the same organisms. Such data (frequently used in survey research) are important in controlling for various threats to the validity of repeated measurement designs (see following for details).

[3] Baltes (1968) conceptualized cross-sectional and longitudinal sequences as both sampling and analysis designs and disputed the validity of the analysis designs suggested in Schaie's general developmental model. This apparent disagreement was reconciled by Schaie and Baltes (1975).

ask: How would the assumption violation of no T effects be reflected in the CS analysis? Although the specific perturbations depend upon the particular data matrix and the sources of effects in that matrix, confounded T effects will generally affect both (1) the estimates of A, reflected in an analysis of variance design as the main effect for A and the C by A interaction and (2) the estimates of the C effects, since the different cohorts will be sampled at different time periods. In short all estimated effects are likely to be perturbed. A simple confounded main effect for T would most likely be reflected in a significant C by A interacton (Schaie, 1973), for (given a confounded T effect) we would be likely to discover that the cohorts differed in maturational pattern simply because they were sampled at different time periods.

This conclusion leads to an interpretational paradox: We cannot distinguish a true C by A interaction from a confounded T main effect, once a C by A interaction has been obtained in the CS analysis, unless we have a *strong theory* which not only hypothesizes a true A by C interaction but also specifies (1) the cohorts over which the interaction holds and (2) the direction of the A effect in each cohort. When such specification is possible, we may examine the pattern of effects to determine if the obtained C by A interaction matches our theoretical specification; if it does not, then some time-related confound is relatively more likely. While a significant C by A interaction which is inconsistent with, or not predicted by, our theory may indicate the presence of a time-related confound, it could also be a reflection of a true C by A interaction which is not of the form specified by our model. The absence of a C by A interaction is not sufficient evidence to conclude that no confounded T effects exist; the power of our test of interaction may not detect its presence, or the effects might be localized to a small subset of occasions, in which case our estimates of A, C, and A by C effects will be biased. This is the essence of the interpretational indeterminacy in sequential analysis: If the assumptions which justify the design are violated, the effect estimates obtained are to some degree inaccurate. Given a strong theory about the nature and direction of estimated and confounded effects, the interpretational problem may be reduced to estimating the relative likelihood of confounded T effects; given the pattern of effect estimates, but in the absence of strong theory, the meaning of the pattern of results from the CS (or any other sequential design) may remain obscure.

One of the positive implications of the relationship of strong theory to the interpretation of sequential designs is that the theory may sufficiently specify the pattern of effects so that an invalid design (i.e., one in which the major assumption is violated) may actually provide useful information about the confounded effect. Consider the cross-sequential [XS] design, which crosses C and T under the assumption of no A effects. A developmental psychologist might well ask: Why should I estimate such a design if effects of A are my primary interest? There are two points to be made in answering this question: (1) The XS design may be applied when longitudinal data are available only for a limited number of measurement occasions (time periods) but a wide range of cohort groups and (2) given a strong developmental theory about the nature of the confounded A effects, a misspecified XS design (in the sense that A effects are nonzero) may provide valuable information about the significance of A effects in both the T and the C by T effects. With regard to the first point, the XS design is feasible with only two measurement occasions, while a CS design requires at least three. The number of measurement occasions required to estimate CS designs which span a relatively large age and/or cohort range can be prohibitive were we to insist that no analysis of the data should be performed until the CS design appropriate to the research question was possible.

To illustrate this point consider again Figure 7-1. A CS design following three cohorts (1880 to 1900, say) over an equivalent 20-year age range (10 to 30, say) would require sampling from 1890 through 1930; an XS design following subjects longitudinally for 20 years would initially be possible in 1910. The CS design is still the method of choice, in that it provides convergent information on the age span 10 to 30 over three separate cohorts, but the misspecified XS design provides some information about development in that age span which, accompanied by theoretical notions about the developmental phenomenon, may be used to make preliminary inferences on the pattern of age effects as represented (confounded) in the T and C by T parameters. Schaie's early work on the sequential analysis of intelligence began by assuming misspecification in an XS design and attempting to draw preliminary inferences regarding C and A effects (cf. Schaie, Labouvie, & Buech, 1973; Schaie & Strother, 1968). It is always preferable to estimate the "true" parameter effects from the appropriate design—i.e., one which makes the correct limiting assumptions; however, it will often be the case that the developmental psychologist must settle for something less than the optimal design, if only temporarily.

Schaie's "Most Efficient Design." Once we allow for a stepwise approximation to the CS design as measurement occasions are added to the data matrix, the question arises on the best way to sample from the age-by-cohort matrix. Most investigators who wish to ask questions with respect to both IAC and IEV nevertheless would like to limit the time course of their longitudinal sampling. In addition theoretical notions about different developmental models applying differentially to subsets of variables in a multivariate set of measures may suggest the need to apply alternative sequential designs to different subsets of dependent variables. Schaie (1965) initially proposed a sequential *sampling* design termed the "most efficient design," which maximizes the potential design applications, given that sampling must begin at some occasion, defined as Time 1. It is as follows:

1. Draw a random sample from each cohort within the age range of interest and measure at Time 1. (Score 11)
2. Obtain a second measurement on as many subjects as possible who were initially tested at Time 1 at Time 2. (Score 12)
3. Draw a new random sample from each cohort tested at Time 1 plus a sample from the next younger cohort and measure at Time 2. (Score 22)
4. Get a third measurement on as many subjects as possible that were measured at Time 1 and Time 2 at Time 3. (Score 13)
5. Obtain a second measurement on as many subjects as

possible who were first tested at Time 2 at Time 3. (Score 23)

6. Draw a third random sample from each cohort tested at Time 2 plus a sample from the next younger cohort and measure at Time 3. (Score 33)

Note that Scores 11, 22, and 33 provide a cross-sectional sequence while Scores 11, 12, and 13, or 22 and 23, will provide longitudinal sequences. Given such data collection it is possible to examine the cohort-sequential model for each set of two cohorts (Scores 11, 12, 22, and 23) or to examine the cross-sequential model for two replications (Scores 11 and 12 as well as 22 and 23). Scores 13 and 33 will permit controls for practice and experimental mortality (see following and Schaie, 1972b, 1977).

Remaining Confounds in Longitudinal Sequences: Possible Solutions

Confounds Not Directly Resolved by the Cohort-Sequential Strategy. In their pure form the previously described approaches will aid the developmental psychologist to estimate maturational effects while controlling for confounds due to history and certain simple selection effects. Other threats to validity of developmental studies can often be controlled by further design refinements.

Experimental mortality. As discussed briefly before, human panels are rarely maintained in their entirety during a longitudinal study. One must check therefore whether subject attrition has been random or systematic with respect to the dependent variables. The most straightforward approach is upon completion of the first follow-up test to segregate Time 1 scores into those for subjects who were successfully retested at Time 2 and those who failed to reappear. Although most investigators of this issue have found that dropouts are, on the average, less able and have different personality characteristics, it does not follow that dropouts will have different age patterns than do the retest survivors (e.g., Schaie, Labouvie, & Barrett, 1973). It does not follow either that systematic dropouts will be maintained upon subsequent retests (Gribbin & Schaie, 1978). Whether or not attrition is subject to secular trends or to cohort effects can be assessed by suitable modifications of the simple sequential models (cf. Schaie, 1977; Schaie & Parham, 1974). The most straightforward control for experimental mortality involves the comparison of cross-sectional sequences (e.g., following successive samples tested only once from the same cohort such as Scores 11, 22, and 33). The problem here is in increased sampling variation and the inability to consider IEV in IAC.

Testing effects. When ability tests are given repeatedly, it is possible that substantial practice effects occur (e.g., Hofland, Willis, & Baltes, 1981). Also the administration of attitude scales or personality tests may tend to produce modification of attitudes or social desirability values of questionnaire items. A direct test of practice effects is possible by comparing performance of individuals of the same age at the same point in time but who differ in level of practice (e.g., comparison of Scores 12 and 22). Interaction of practice with age, cohort, or time-of-measurement effects can be studied, as well as the possible interaction of practice and experimental mortality. All such designs, however, require addition of further follow-up data collection and a combination of cross-sectional and longitudinal sequences (cf. Schaie, 1977, for details).

Changes in instrumentation. The need to maintain the same methods of data-collection reduction and analysis across measurement occasions is obvious, since changes in instrumentation will introduce time-of-measurement effects in the data. An additional instrumentation problem in longitudinal studies conducted over extensive periods of time is that it may become unavoidable to change part or all of the assessment battery. This may be the case because tests given to subjects when they were children may no longer be valid for the same subjects as adults. Batteries may also require change when tests shift in validity due to cultural change affecting their construct validity (cf. Gribbin & Schaie, 1977). In such cases control samples may be needed to whom both old and new instruments have been administered. Alternately it may be possible to compare factor scores upon the application of appropriate techniques of comparative factor analysis (Jöreskog & Sörbom, 1979).

Statistical regression. Observed changes in level in longitudinal studies may be no more than consequences of insufficient reliability of measurement instruments. Particularly in the case where comparison of several levels of ability or standing on other classificatory dimensions is sought, regression effects need to be examined. A general discussion of this problem is provided by Furby (1973), and a method proposed by Baltes, Nesselroade, Schaie, & Labouvie (1972) may be useful in assessing the extent of the problem. In that method, bottom and top scores are divided at Time 1 and compared at Time 2; if performance gradients are not parallel, bottom and top scorers are then divided at Time 2 and compared at Time 1. Regression effects are demonstrated if the gradients obtained under the two methods show opposite direction.

True time-of-measurement effects. When secular trends are expected as in short-term studies, it may be reasonable to switch to one of the alternate strategies deduced from the general developmental model. For example for periods of the life span where little developmental change is expected, the *XS* model might then be appropriate. Likewise where cohorts are defined as narrow bands with little likelihood of substantial cohorts shifts, the *TS* model may then be reasonable to test for time-of-measurement effects as contrasted to age differences (see Schaie & Parham, 1974).

Sampling bias and volunteer behavior. Here only collateral studies or knowledge of the relation between parent population and obtained samples will help. One such collateral study might be to investigate volunteer behavior under alternate conditions both with respect to rate of responding and performance on the dependent variables of interest. An example of such a collateral study has been reported by Gribbin and Schaie (1976) as to the effect of offering or not offering a monetary incentive. In the latter study no differences were found in rate of volunteering or performance in intelligence tests, but some personality questionnaire differences were noted.

Effects of changing populations. Secular changes in the demographic composition of a population present difficulties for obtaining comparable samples in sequential designs. Any changes over time in population characteristics will tend to produce samples which cannot be assumed to differ only in age, birth year, or time tested. If the population is changing, it may be desirable to shift a sampling without replacement to a sampling with replacement model (e.g., Gribbin, Stone, & Schaie, 1976). It may even be appropriate to attempt to match samples on certain characteristics (although matching often creates as many problems as it solves). In any case, it becomes most important in sequential sampling for the researcher to have good demographic information about the population from which a panel is drawn, as well as information about shifts in population characteristics over the time frame of the longitudinal study.

Additional Design Considerations

In addition to attending to the confounds inherent in designs analyzing longitudinal and cross-sectional data matrices, the developmental psychologist must consider several other issues relevant to selection of a sequential design for a given research application. Several of these theoretical and practical considerations are discussed in this section.

Unequal Sampling Intervals. A potential source of problems for data analysis and interpretation in sequential designs is the use of unequal sampling intervals over time, where disproportionate numbers of years intervene between occasions of measurement. An example would be a study in which the initial measurement occasion was, say, 1980, and subsequent testings occurred in 1985, 1986, 1990, and 1995. These unequal intervals of measurement require special analytic techniques if age is to be treated as an interval scale in estimating aging or growth-curve parameters (see following); moreover, the unequal intervals will have the effect of "deorthogonalizing" the age or time factor with the cohort factor. Unequal intervals of measurement do not produce insurmountable analysis problems, but they increase the complexity of the analysis procedure and should be avoided in the name of parsimony unless special considerations relevant to the problem area require them.

Unequal Factor Intervals. Another design problem in sequential strategies arises when the two factors included in the design involve different time spans. This issue is not isomorphic with the unequal sampling intervals problem, for here we refer to different numbers of levels of each factor (be it *A*, *C*, or *T*) even when the time interval for each of the factors is held constant. Figure 7-2 illustrates the distinction; in the lower panel (Figure 7-2B) the unequal factor interval problem arises because more levels of *C* are measured than are levels of *T*. This type of unequal interval will be quite common in sequential sampling, because it is always possible to sample a wide range of birth cohorts at a single point in time, and replicate sampling over time would result in a matrix which contains a wider cohort range than a time range (at least) until the time span (in years) between first and most recent measurement occasions equaled the time span (in years of birth) between the most recent and

A. An equal time interval CS design (3 cohorts, 3 ages, 20 year span)

COHORT	10	20	30	40	50
1880	1890	1900	1910	1920	1930
1890	1900	1910	1920	1930	1940
1900	1910	1920	1930	1940	1950
1910	1920	1930	1940	1950	1960
1920	1930	1940	1950	1960	1970
1930	1940	1950	1960	1970	1980

Age

B. An unequal time interval CS design (4 cohorts, 2 ages, 30 year cohort span, 10 year age span)

COHORT	10	20	30	40	50
1880	1890	1900	1910	1920	1930
1890	1900	1910	1920	1930	1940
1900	1910	1920	1930	1940	1950
1910	1920	1930	1940	1950	1960
1920	1930	1940	1950	1960	1970
1930	1940	1950	1960	1970	1980

Age

Figure 7-2 Schematic illustrating the unequal time interval problem for a cohort-sequential design

most remote cohorts initially sampled. Thus a cross-sequential analysis after two times of measurement will probably contain disparate time spans, with more cohorts than times of measurement in the design. In fact the cross-sequential analyses of intelligence by Schaie and co-workers (e.g., Schaie, Labouvie, & Buech, 1973; Schaie & Labouvie-Vief, 1974; Schaie & Strother, 1968) involve analyses of such disparate time spans.

Unequal time spans produce a type of "bias" in sequential designs—namely, that the expected value of the variance components for each factor will be unequal and in direct proportion to the ratio of the different time spans (Adam, 1977; Botwinick & Arenberg, 1976). Thus the differences in time spans in the factors will be reflected in an analysis of variance as in differences in the size of omnibus F-ratios (that is, F-ratio testing the null hypothesis of equivalence of all the marginal means for the factors) or in different values for proportion of variance estimates such as the intraclass correlation or ω^2. If the cohort factor is sampled over 20 years of birth and the time factor is sampled over five years of measurement, then we would expect that the omnibus F-test testing the hypothesis of no cohort effects would be roughly four times larger than the omnibus F-test for no time effects even if the two factors had roughly the same magnitude of effect for each unit of time on our interval scales (i.e., years).

This effect is a natural consequence of the unequal time intervals and is not a source of bias in the statistical sense of bias in estimators. The omnibus F-test in both cases is the proper test of the null hypothesis of no effects across all levels of the factors involved; as the number of levels of the factor increases, the likelihood also increases that at least two of the subclass means are reliably different. Thus the unequal time intervals do not "bias" the results and do not in any way limit the investigator to using equal time spans in sequential designs. The primary effect of unequal intervals is that the magnitude of F-ratios or proportion of variance estimates cannot be used in a direct comparison of the relative importance of the two factors under study. Using our previous example it would be erroneous to claim that a cohort effect is more important than the time effect simply because the proportion of variance for cohort was greater, since that pattern would be predicted by the unequal time intervals alone. A better test of the hypothesis would be that the ratio of variance accounted for exceeds a level reliably greater than that expected by the disproportionate time spans alone.[4] In the presence of unequal time spans, the investigator should be careful not to make mistaken inferences on the relative importance of the two effects being estimated; however, there need be no worry that the significance tests are invalidated by the unequal intervals.

Age, Cohort, and Time as Continuous Variables. The argument could be made that descriptive inferences using the time-related variables age, cohort, and time are less than maximally powerful unless these variables are treated as continuous and not as categorical variables. Under this argument the sequential designs discussed above should not arbitrarily categorize these variables and use analysis of variance but treat them as continuous and use multiple regression (e.g., Buss, 1979–1980). Since multiple regression and analysis of variance are functionally isomorphic (Cohen & Cohen, 1975; Kerlinger & Pedhazur, 1973) the question is not one of appropriate analysis technique but whether too much information is lost by pooling the continuous time intervals into discrete categories. The loss of information obviously depends on the size of the categorical intervals pooled and the within-category covariation thereby ignored. In general we suspect that the regression approach is not likely to yield vastly different inferences from the analysis of variance approach in practice as long as sample sizes are not small (say not less than ten per cell) and the time intervals defining age or cohort are not large (say, less than ten years). Part of the reason for our conclusion is that chronological age per se is an imperfect measure of biological, psychological, and social aging (Wohlwill, 1973); hence, pooling over small intervals is not likely to lose a great deal of predictive power, given the measurement error inherent in the chronological age variable. We recognize that regression approach has probably been underused in sequential data analysis, primarily because the sampling procedures (measurement at discrete time intervals) used in longitudinal research led directly to the use of categorical analysis (i.e., analysis of variance) methods.

Practical Issues. The preceding discussion leads directly into consideration of the following questions: (1) How large a time interval should exist between measurement occasions? (2) How wide a range of ages and/or cohorts should be sampled? and (3) How large should the sample size be? General answers to these questions cannot be given since they depend directly on the content area of interest, the developmental hypothesis to be tested, and the level of prior knowledge about the phenomenon under study. Consideration of the length of time interval in longitudinal sampling depends upon the tradeoff between the need to measure change in the minimum time possible and the concern that repeated testing in short periods of time will greatly increase the probability of unwanted practice effects. Consideration of the size of the age or cohort range to be sampled similarly depends upon the hypotheses the investigator has about the critical developmental periods and the range of birth years over which generational differences are expected. In general it is probably desirable to include a wider range of ages and/or cohorts than are hypothesized to show differences, particularly in an initial exploratory study, so that the hypothesis may be falsified with respect to certain ages and/or cohorts; however, the inclusion of additonal levels of cohort and age must be weighed against the cost of including additional subjects and measurement occasions in the design. With respect to sample size, the investigator may wish to perform a power analysis (Cohen, 1977) to predict how many subjects will be required to yield statistical significance for a lower bound estimate of effect size (assuming one is

[4] Botwinick and Arenberg (1976) criticized Schaie and coworkers for direct comparisons in cross-sequential analysis using disparate time spans. Their criticism was recognized by Schaie and Parham (1977) as valid (with respect to the disparate time spans only). The latter paper reported on equal time-span cohort-sequential analyses, in which unweighted comparison of F ratios was justified.

available); however, in longitudinal sampling one should allow for a rate of experimental mortality in developing this estimate. The attrition rate will depend upon the nature of the population and the sampling procedure (e.g., Rosenthal & Rosnow, 1975); in Schaie's studies of unpaid volunteers from a prepaid health plan, sampled at seven-year intervals, attrition rate in the longitudinal sequences has been roughly 40 percent (Schaie, 1979).

Decision Rules for Age, Cohort, and Time Effects. As discussed before, one of the problems in implementing a sequential design approach is the consequence of model misspecification and the difficulty in drawing valid inferences on the presence of age, cohort, and period effects in the absence of strong theoretical posture on the likelihood of these effects being present in the data. Originally Schaie (1965) attempted to formulate some decision rules for deciding on the presence or absence of the three effects by comparing the results from different sequential analyses of the same data matrix, a procedure criticized by Baltes (1968) and Buss (1973), among others. Schaie's decision rules were based upon an intuitive rationale for teasing apart effects which, quite frankly, has been shown to be of questionable validity—and these decision rules should no longer be taken seriously. The critical problem is that in the presence of effects for all three factors, and in the concurrent absence of strong theoretical specification of the pattern and magnitude of some of the effects, there is at present (and in the foreseeable future) no method available for estimating all the three effects and their interactions simultaneously. The decision rules by Schaie (1965) will not generally lead to the "true" model—that is, the one which produces unbiased estimates of all the effects operating in a given domain of study. An *atheoretical* exploratory study attempting to estimate effects for all levels of age, cohort, and period and their joint interactions is unlikely to be a fruitful enterprise.[5]

Additional Between-Subjects Factors. The investigator may wish to partition the between-subjects portion of the sequential design according to additional individual difference variables of interest. This partition may be either a priori (that is, the factor is explicitly included in the sampling design) or a posteriori (the partition is based upon individual differences determined during or after the first measurement occasion). Examples of such grouping factors are sex, experimental mortality (Schaie, 1977), or health status (Hertzog, Schaie, & Gribbin, 1978). In some cases (e.g., experimental mortality) the partition must be a posteriori since it is based upon events occurring after the first measurement occasion. The major problem with a posteriori group definitions is that they will usually result in a loss of statistical power caused by the reduction in subgroup size. It is therefore preferable, wherever possible, to include additional grouping factors at the time of initial design so that sampling may be done with respect to all subgroups.

The investigator should be aware that the inclusion of additional groups also incurs the risk of additional sampling by treatment interactions, which may have ramifications for the age effects being estimated. It might be the case, for example, that middle-aged women and middle-aged men would be differentially representative of their respective populations due to differential availability for sampling (determined obviously by the type of sampling procedure employed).

Alternative Designs for Sequential Data Analysis

Simultaneous Estimation of A, C, and T. An important alternative type of sequential designs eschews Schaie's standard sequential designs in an attempt to simultaneously estimate components for all three factors, *A*, *C*, and *T*, in a single design. These designs operate on the same sequential data matrices (i.e., cross-sectional and longitudinal sequences) but do not use traditional analysis of variance designs or analysis techniques to estimate effects of only two of the three factors, as in the *XS*, *TS*, and *CS* designs. We refer to these alternative models as *additive effects models* since the basic procedure is to postulate no interaction components involving any of the three factors, *A*, *C*, and *T*. Under this additivity assumption, it is possible to simultaneously estimate parameters for some of the *A*, *C*, and *T* effects if enough suitable assumptions (in the form of restrictions on the effects that are present in the analysis) are imposed to make all remaining parameters estimable.

The approach was first discussed by Mason, Mason, Winsborough, and Poole (1973) and later modified and advocated by Donaldson (1979) and Horn and McArdle (1980). Consider a cross-sectional data sequence of the type shown in Figure 7-1. Under the additivity assumption, the cell means in the population may be represented by the following equation:

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k \qquad (1)$$

where α refers to effects for age i, β to effects for cohort j, and γ to effects for time k. For any individual, the linear model will also contain an individual differences error component:

$$y_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijkm} \qquad (2)$$

In normal ANOVA applications, we might impose side conditions of the form $\Sigma_{\alpha_i} = 0$, etc., in order to uniquely define the parameters, and to enable the usual ANOVA tests of the null hypotheses:

$$H_\alpha\text{: all } \alpha_i = 0;\ H_\beta\text{: all } \beta_j = 0;\ H_\gamma\text{: all } \gamma_k = 0 \qquad (3)$$

[5] In her critique of sequential strategies and Schaie's (1965) decision rules, Adam (1978) demonstrated that the expected value for *C* in a 2 by 2 cross-sequential analysis was equivalent to the expected value for *A* in a 2 by 2 time-sequential analysis involving the same cells of a cross-sectional sequence. Adam's point is well taken, although it depends upon the assumption that all *A*, *C*, and *T* effects are nonzero, and it is limited to the 2 by 2 case. However Adam goes well beyond this demonstration to infer that it somehow validates criticisms of Schaie's work by Horn and Donaldson (1976) and that it would be preferable to use some two-factor sequential design in exploratory situations. Adam is incorrect on both points; in the latter case her assumption that all *A*, *C*, and *T* effects are nonzero insures that no sequential design will produce unbiased estimates of the various effects. Since Adam does not specify her two-factor design, we cannot be more explicit as to the confounds inherent in the two-factor design she seems to prefer. Her model is explicitly additive, and thus she would be able to estimate all but two of the *A*, *C*, and *T* effects under the additive effects model we describe later (although violation of the additivity assumption would also produce biased estimates).

However, given the linear dependency among the three factors, A, C, and T, no solution exists for all $\alpha_i\beta_j$, and γ_k. Mason, Mason, Winsborough, and Poole (1973) showed, however, that if an a priori assumption could be made regarding the equality of two parameters (e.g., $\alpha_1 = \alpha_2$), this assumption would be sufficient to remove the indeterminacy and to just identify the other parameters. Then statistical estimates for these parameters could be found. They also pointed out that additional assumptions of the same type would place further restrictions on the model leading to overidentification of the remaining parameters (which has desirable properties in statistical hypothesis testing of the models). The parameter estimates in such a procedure are, of course, completely dependent upon the accuracy of the equality assumptions used to enable the estimation procedure; different assumptions would produce different estimates for the effects.

Donaldson (1979) extended the Mason approach by treating it under the framework of full rank linear modeling (Searle, 1971; Timm & Carlson, 1975). Horn and McCardle (1980) further generalized the approach to restricted modeling of mean and covariance structures. In the process of extolling the virtues of the additive effects models, these authors have been highly critical of both Schaie's sequential designs and traditional ANOVA applications to analyze them, stating explicitly that the model testing approach is superior to the "traditional" sequential strategies approach. It has become important to consider the advantages and disadvantages of the simultaneous additive effects model by itself and in comparison to sequential strategies; if the additive effects approach could be considered invariably superior to the traditional sequential designs of Schaie, it should obviously supplant them as the method of choice.

Before proceeding with an evaluation of the additive effects model, let us recognize that there are two separate criticisms inherent in the Donaldson-Horn-McArdle critique: (1) that the parametric model of Schaie's sequential approach is invalid and (2) that the statistical analysis procedures employed by Schaie in his empirical applications of sequential strategies are invalid. With regard to the second point, the critique is pointed toward the application of standard ANOVA techniques. There is no question that the ANOVA procedures used by Schaie (e.g., Schaie & Labouvie-Vief, 1974; Schaie and Strother, 1968)—namely, use of unweighted means, univariate ANOVA on multiple dependent measures—are dated by more sophisticated multivariate approaches (of the kind we outline later). Moreover it is certainly true that Schaie and colleagues' previous application of traditional ANOVA techniques emphasized interpretation of omnibus F-tests and did not focus on estimation of individual effects (i.e., the αs etc.] or on a priori or a posteriori contrasts among cell means in order to delineate more precisely the source of the significant differences. Comparisons among individual-cell means were done on the basis of longitudinal or cross-sectional gradients across observed means. The use of classic univariate ANOVA models is no longer the staple of Schaie's statistical analyses; in any case it should not be the focus of a discussion of the merits of the parametric models of the additive effects approach. Having now confessed past "sins," we turn to the issue at hand!

In our view the additive effects model represents an important contribution to the area of sequential methodology, but it is just as flawed and imperfect as Schaie's traditional sequential designs, for it is no more or less valid than the assumptions invoked to enable estimation of the A, C, and T effects. In the case of the additive effects model, there are two sets of assumptions: (1) All interaction effects are zero and (2) at least two, and possibly more, main effect contrasts are equal. Donaldson (1979) and Horn and McArdle (1980) are both quite explicit on the dependence of the estimates of the additive effects model on the validity of the assumptions of the second type, but they completely ignore the importance (in our view, the more critical importance) of the validity of the assumptions of the first type. Similarly the validity of the *XS*, *TS*, and *CS* designs is contingent upon the assumption that the unanalyzed components are all zero. Both the traditional sequential strategies and the additive effects model suffer from a common problem: Given invalid assumptions, we obtain inappropriately biased and invalid effect estimates.

We may formalize the problem in the following way. Given a fixed range of ages, cohorts, and periods of measurement, a general linear model for the means in the population is:

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \quad (4)$$

where the joint terms (e.g., $\alpha\beta$) denote interactions. For a given population and construct(s), some or all of the effects may be zero, but the preceding equation describes the general case in which all effects are nonzero. In the equation we are indicating true population parameters, not statistical estimates of those parameters. The problem is that, although the effects for all the terms in Equation 4 may be present in the population (all are in principle theoretically distinct), the linear dependency among the measures of A, C, and T makes it impossible to estimate all the effects independently. In fact the linear dependency precludes one from ever estimating the three-way interaction among A, C, and T, even though it might be theoretically meaningful. If one wishes to estimate all the possible effects for any given two-way interaction, say $(\alpha\beta)_{ij}$, then none of the other effects for either of the remaining two-way interactions is estimable. If one wishes to estimate all the effects for two of the main effects, then none of the effects for the remaining factor is estimable. In all cases the design is limited by the degrees of freedom contained in the sampling design, which is of "deficient rank" with respect to A, C, T because of the linear dependence among the factors.

Given this state of affairs, the investigator may only obtain valid effect estimates for some of the population parameters under the assumption that the parameters not estimated are in fact zero in the population. An *XS* design assumes all terms involving α_i—i.e., α_i, $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\alpha\beta\gamma)_{ijk}$—to be zero; a *CS* design assumes all terms involving γ_k to be zero. The additive effects model assumes all interactions—i.e., $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, and $(\alpha\beta\gamma)_{ijk}$—to be zero and imposes at least one additional assumption on the main effects (of the form $\alpha_1 + \alpha_2$ etc.) in order to estimate the remaining effects. When all is said and done, none of these models is applicable to all

developmental problems. Theorists are likely to differ on the merits of any of these models to a given problem; Glenn (1976) and Baltes, Cornelius, and Nesselroade (1979) doubt the usefulness of the additive effects model because they suspect that the hypothesis of no interactions is rarely, if ever, likely to be true, while Donaldson (1979) and Horn and McArdle (1980) question the validity of sequential strategies because they doubt that all effects attributable to one of the *A*, *C*, and *T* factors are ever likely to be zero in the population. This is the dilemma of descriptive research designs involving parametric treatment of *A*, *C*, and *T* effects; we must make limiting assumptions to estimate any of the effects, and these assumptions must be theoretically sound for the estimates to have maximal utility. As pointed out by Baltes et al. (1979), there just is no purely statistical solution to the problem.

Before concluding this section we should note another potential problem with the additive effects model. The simultaneous estimation of *A*, *C*, and *T* effects might mislead one into assuming that meaningful estimates of all *A* and *C* effects could be obtained from simply taking two cross-sectional samples (i.e., the smallest possible cross-sectional sequence). In fact a just-identified solution attempting to estimate all *A* and *C* effects (except two which are set equal) is unlikely to lead to useful estimates because of the presence of a high degree of nonorthogonality among the effect contrasts used in the statistical analysis. There will be a powerful suppressor effect operating if all the estimates are made simultaneously. If the investigator is interested in estimating *A* and *C* effects, there is no better matrix of observations than the age-by-cohort matrix discussed previously in the context of *CS* designs. As the cross-sectional sequence increases in measurement occasions, then an additive effects design estimating *A* and *C* effects will experience a decreasing problem with suppression with respect to *A* and *C* effects (the same principle applies to longitudinal sequences). Thus it is not the case that the additive effects model obviates the need for extended sequential sampling.

As stated before we believe the additive effects model to be a significant contribution to the area of sequential methodology, not because it should supplant other sequential designs but because its proponents have demonstrated how the sequential designs of Schaie are only one way of partitioning a sequential sampling matrix. Given theoretical justification other types of designs may also be formulated. Indeed one possibility (which we have not thought through in any detail) is that, under a theoretically sound set of restrictions on different *A*, *C*, and *T* effects, it may be possible to also estimate some partial interaction effects (cf. Boik, Note 1) in the same design. The fact that individual effects may be explicitly represented in linear models by a set of vectors of contrast coefficients describing relations among the cell means increases the investigator's flexibility in matching theory to design and estimation.

Replacement of Age, Cohort, and Time

Since age, cohort, and time actually represent marker variables (Wohlwill, 1973; see preceding) for underlying causal processes, it is undoubtedly the goal of developmental explanation to replace these variables with the process-oriented variables thought to actually determine the *A*, *C*, and *T* effects. As pointed out by several authors (e.g., Baltes et al., 1979; Buss, 1979–1980; Schaie, 1977–1978), such replacement requires a valid theory for the source of the underlying effects and a suitable method for measuring the processes under study.

The advantage of study of the "real meaning" of cohort membership (Baltes et al., 1979; Rosow, 1978) or age from a strictly descriptive viewpoint is that it implicitly removes the linear dependency among *A*, *C*, and *T* by replacing *A* or *C* with process variables which are applicable to all levels of the other two factors (see Mason et al., 1976). One could then proceed to estimate descriptive parameters for *A*, *C*, and *T* effects without the constraints discussed in the previous section.

The ultimate goal, however, would be to replace all three factors with process variables and to account for change over time in behavior on the basis of explicit knowledge of antecedent-consequence relationships (Baltes & Willis, 1977, 1979). The ideal method for proceeding from construct and variable definition to statistical estimation in such causal models is the use of structural regression (Baltes, Reese, & Nesselroade, 1977; Buss, 1979–1980; Rogosa, 1979) as outlined by many behavioral scientists and mathematical statisticians (e.g., Duncan, 1975; Goldberger & Duncan, 1973; Heise, 1975; Jöreskog, 1973). Structural regression methods provide the most comprehensive means by which causal influences may be modeled among nonmanipulable individual-differences variables of the type inherent in developmental research. A detailed examination of the issues inherent in causal modeling is beyond the scope of this chapter; the reader is referred to the references just cited. Further discussion of structural regression approaches in the context of developmental analysis is given in the following sections.

STATISTICAL METHODS FOR DEVELOPMENTAL ANALYSIS

Developmental Hypotheses about Means

The majority of developmental studies are interested in testing hypotheses about change in level of performance over time. When the hypothesis involves developmental change in the population, interest focuses on change in the population means with age, which may be summarized in an average growth curve. With respect to the population means, the developmental psychologist wishes to know (1) the direction of change in mean levels with age (time) and (2) the magnitude of developmental change, expressed in the unit of measurement of the interval scale *X*. Developmental patterns in performance level could be multidirectional (nonmonotonic), monotonically increasing, monotonically decreasing, or stable with increasing age. Knowledge of the direction of developmental change will rarely suffice, however; usually the investigator requires an estimate of the magnitude of such change.

The sufficiency of any set of summary statistics describing the average developmental function is depen-

dent upon the complexity of that function. When the change in means is linearly increasing or decreasing over time, the magnitude and direction of change may be economically expressed as the slope of the linear function—i.e., change in X per unit time (age). When the developmental function is nonlinear, description of direction and magnitude of change is more complex. When the developmental function may be specified, the summary statistics derive from the parameters of a fitted function (e.g., a Gompertz curve). Given a nonlinear developmental function whose exact form is left unspecified, the direction and magnitude of change may be represented in any set of summary statistics which encapsulate the mean difference between ages. Two common sets of statistics are mean contrasts among adjacent ages (occasions) or the regression weights from an orthogonal polynomial equation.

In the multivariate case the investigator would be interested in the consistency of the developmental function across variates—that is, the extent to which the means vary in direction and magnitude of change across different measures. When the data have been collected for multiple groups from the population, the investigator's interest focuses on the consistency of the developmental function across groups. Different groups might have coincident functions (a single developmental function in common, such that the curves lie on top of one another), parallel functions (equivalent changes with age but constant mean differences between groups at each age [occasion]), or divergent functions (group differences in developmental change, with or without group differences in means at the initial age measured).

Developmental Hypotheses about Covariance Matrices

Although most developmental studies have focused on developmental changes in mean performance levels, developmental hypotheses about changes in the range and ordering of individual differences with age may be tested by examining the appropriate elements of the covariance matrix of the observations. The variance parameter reflects the magnitude of the dispersion of individuals around the population mean; thus changes in variances with age indicate age change in interindividual variability (IEV). The covariance elements among repeated measures of a single variable reflect the extent to which the ordering of individuals about the means is consistent across measurement occasions. If individuals maintained fixed positions relative to the mean, the covariance between the measure at any two occasions would equal the product of the variances, and the correlation between the two measures would be 1. Thus the covariances reflect the stability of IEV with age.

The interpretation of developmental changes in variance and covariance parameters with age depends in part upon the ordering of such changes. A systematic increase in variability over time might well imply IEV in intraindividual change (IAC) since the increasing dispersion of individuals about the average developmental function might reflect divergence of individual developmental functions from the average developmental function. Many complex developmental hypotheses may be modeled explicitly in terms of their implications for the covariance structure of the measurement variable over occasions.

For the multivariate case the investigator may be interested in the consistency of developmental changes in variance and covariance elements across multiple measures, which would be reflected in the similarity of changes in variance and covariance elements over occasions. If the design involves measurement of multiple subgroups from the population, interest will also focus on the consistency of the changes on variance and covariance elements between the different groups. As with group comparisons on means, one can ask whether the groups show equivalent changes in individual differences at each point of the developmental function.

Statistical Tests of Hypotheses concerning Means

As discussed previously most developmental research questions focus on age changes in the means of a group of subjects over time. Hypotheses regarding age changes in mean levels have usually been tested by means of classical analysis of variance (ANOVA), as have most hypotheses in recent psychological research. Increasingly, however, methods of multivariate regression have begun to supplant traditional ANOVA as the statistical approach, fueled by the increasing awareness that the ANOVA and regression approaches are basically the same, ANOVA being a special case of regression with categorical independent variables and an orthogonal experimental design (Cohen & Cohen, 1975). Indeed most statistical packages which now perform ANOVA actually use regression as the computational technique. One of the major advantages of using multivariate regression to analyze ANOVA designs (i.e., categorical sampling designs) is that it is particularly appropriate for the analysis of nonorthogonal sampling designs, wherein the cell sizes are unequal. In sequential data where orthogonality with respect to the subclass subject frequencies is rarely, if ever, obtained, the generality of the regression method for testing hypotheses about the subclass means makes it the method of choice.

The General Multivariate Regression Model. The model for multivariate regression (also known as the general linear model) has been extensively treated in a number of texts (e.g., Bock, 1975; Searle, 1971; Timm, 1975). The general multivariate model for an individual in the population is:

$$\underset{\sim}{y} = \underset{\sim}{x} \cdot \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{5}$$

where $\underset{\sim}{y}$ is a $p \times 1$ vector of dependent variables, $\underset{\sim}{x}$ is a $q \times 1$ vector of independent (predictor) variables, $\underset{\sim}{\beta}$ is a $p \times q$ matrix of regression coefficients, and $\underset{\sim}{\varepsilon}$ is a $p \times 1$ vector of error components.

The interpretation of the regression coefficients in $\underset{\sim}{\beta}$ depends upon how the independent variables are structured in $\underset{\sim}{x}$. If, as in the present case, we are concerned with the analysis of categorical sampling designs, the independent variables in $\underset{\sim}{x}$ must be structured to reflect the classifications of the design matrix. Generally the method is to fill $\underset{\sim}{x}$ with any set of independent (not necessarily orthogonal) contrasts among the cells in the design

by using the method of dummy coding (Cohen & Cohen, 1975; Searle, 1971; Timm, 1975). Often this is accomplished by defining $\underline{x}$ as a model matrix, $\underline{A}$, consisting of a matrix of ones and zeroes, indicating one effect of the linear model (i.e., each α_i etc.) in each column of $\underline{A}$. A design matrix of this type is of deficient rank; not all such effects may be estimated. In fact, when the grand mean vector $\underline{\mu}$ is unknown and must be estimated, none of the individual effects are separately identified, only differences among the effects may be estimated (Searle, 1971). The general procedure is to reparameterize $\underline{A}$ and $\underline{B}$ in terms of a new basis matrix $\underline{K}$ and a parameter matrix $\underline{\theta}$, where the elements of $\underline{K}$ represent difference contrasts among the effects of the form $\underline{\theta}=\alpha_1 - \alpha_2$, etc. The basis matrix may be selected on the basis of a priori planned comparisons among the means, in terms of effects specified by hypothesis, in accord with the "usual ANOVA constraints," or by specifying certain effects to be zero; in each case it is necessary to reduce the number of independent linear functions of the effects to the rank of the matrix $\underline{A}$ (which is equivalent to the degrees of freedom in the categorical sampling design).

Once the model is reparameterized to full rank, estimation of the regression coefficients and significance tests for these effects in terms of hypotheses about the means follows. The most common method of estimation is the familiar least squares method. The procedures for multivariate significance testing are too complex to be detailed here (see Bock, 1975 or Timm, 1975); the logic corresponds closely to significance testing procedures for univariate regression analysis of categorical designs (Cohen & Cohen, 1975).

We have to this point ignored the complications introduced by the use of longitudinal (repeated measures) designs. Before considering longitudinal sequences, we should point out that the model as specified is well suited to the analysis of sequential designs using cross-sectional sequences, where all *A, C,* or *T* effects are represented as between-subjects effects. The basis matrix simply reflects the contrasts among the cells for the sequential design selected. A particularly useful set of contrasts for cross-sectional sequences is the set of orthogonal polynomial coefficients; *A* or *T* effects may be represented in terms of a polynomial model of a specified degree (see Bock, 1975, 1979).

The presence of nonorthogonality in the sequential design complicates the analysis considerably. Nonorthogonality among the effects arises from two sources: (1) the subclass frequences (cell sizes) are unequal and (2) the linear contrasts in *K* are not orthogonal. The first source is by far the most common and is the rule rather than the exception in sequential sampling designs. The second source would arise if the contrasts were not orthogonal in the sampling design, as in the case for the additive effects model. In the orthogonal case where the cell sizes are equal and the contrasts mutually independent, the main effect estimates are all orthogonal to one another. The order of entry of effects into the regression equation is then arbitrary and has no effect on the sums of squares partition among the effects. As is well known, however, when the design is nonorthogonal, the expected-mean squares of the main effects are not independent but certain terms involve the sum of effects over the levels of the other factors (Bock, 1975; Searle, 1971). The confounding of the main effects and interactions implies that the order of entry of contrasts in the regression equation affects the sums of squares associated with each effect. Thus the usual hierarchical (stepwise) fitting of effects must consider the consequences of fitting α_i before β_j say.

These issues have been considered in detail by several authors (e.g., Appelbaum & Cramer, 1974; Cramer & Appelbaum, 1980; Herr & Gaebelein, 1978; Overall, Spiegel, & Cohen, 1975), and there is no consensus on how the problem should be handled. One approach is to specify on a priori grounds the order of effects entry into the regression equation. This hierarchical model requires that the precedence of certain effects over others may be specified on theoretical grounds. An alternative approach is to adjust each main effect by entering its effect contrasts as the last set of main effects, eliminating the effects of preceding main effects. This approach is advocated by Searle (1971) and others (e.g., Overall & Spiegel, 1969; Overall et al., 1975) as the most valid, especially when the source of nonorthogonality in the cell sizes is nonrandom with respect to the factors in the design. An alternative approach is to use a simultaneous estimation procedure which includes all effects in the design. This approach, termed the *experimental design model* by Overall and Spiegel (also termed the *standard parametric model,* STP, by Herr and Gaebelein) adjusts all effects by simultaneously eliminating the sums of squares shared by the other effects in the design. This method is not available in most multivariate analysis packages using least squares methods to estimate the linear model. It is the model available in the maximum likelihood program LISREL discussed later.

The problem of nonorthogonality is particularly acute for analyses with cross-sectional sequences, where all effects are between-subjects and selection of the nonorthogonal method may affect partition of the sums of squares among *A, C,* and *T* effects. For analyses of longitudinal sequences, the longitudinal effect (either *A* or *T*) is generally orthogonal to other between-subjects factors (e.g., *C,* or sex), and the sampling design may be partitioned into mutually orthogonal subspaces of between- and within-subjects effects. Then the selection of nonorthogonal analysis methods affects only the between-subjects effects.

Longitudinal Sequences. The analysis of sequential designs for cross-sectional sequences involves traditional applications of MANOVA techniques to between-subjects designs; we will not consider this application further (see Bock, 1975; Timm, 1975). Analyses of longitudinal data are complicated by the presence of the within-subjects factors, which requires special statistical treatment. There are several possible statistical approaches to the analysis of a longitudinal data matrix, including (1) classical mixed model ANOVA (see Winer, 1971), (2) unweighted or weighted MANOVA (Bock, 1979), or (3) analysis of covariance structures (Jöreskog, 1974; Wiley, Schmidt, & Bramble, 1973). The reader will probably be most familiar with the classical mixed model ANOVA, which in fact was used by Schaie and co-workers in many previous analyses of sequential designs. Advances in statistical treatment of longitudinal data over the last decade have badly dated the classical mixed model

ANOVA, however, and we no longer advocate its use except in unusually favorable circumstances. Indeed in the following sections mixed model ANOVA is discussed mainly to provide a background for preferred alternative methods.

General considerations. As discussed before, two sequential designs are possible, given data from a longitudinal sequence: the *XS* design, crossing *C* with *T* (repeated measures on *T*), and the *CS* design, crossing *C* with *A* (repeated measures on *A*). We focus on the analysis of the *CS* design; the generalization to the *XS* design is straightforward. The presence of the repeated measures facet of the *CS* design introduces the random factor *subjects* into the design (nested within *C* groups). Restricting our consideration of the *CS* design for the moment to the univariate case, the linear model for this design is:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \tau_{i(k)} + \alpha\beta_{jk} + \beta\tau_{ij(k)} + \varepsilon_{ijk} \quad (6)$$

where α_j are the $j = (1 ...., _j)$ age effects, β_k are the $k = (1, ..., k)$ cohort effects, $\tau_{i(k)}$ are the effects for the $i = (1, ...i_k)$ subjects, nested in the kth group, and the remaining effects are the associated interactions and the individual error component.

Mixed-model analysis. As is well known the virtue of the repeated measures design is the increased power of the statistical tests due to the removal of the subjects effect from the error term. With more than two levels of the age factor, the conventional mixed model approach requires pooling of the error *SS* over the multiple degree of freedom error subspaces. Unfortunately the assumption of sphericity in the error space necessary for this procedure is often violated and the traditional mixed model significance tests for the age and cohort-by-age interaction will often have inflated Type I error rate (Greenhouse & Geisser, 1959; McCall & Appelbaum, 1973).[6] The two major solutions to the problem are (1) use of corrected *F*-ratios (by adjusting the degrees of freedom) as recommended by Greenhouse and Geisser (1959) or (2) use of a *multivariate* significance test for the *univariate* repeated measures effects (McCall & Appelbaum, 1973). There is some debate as to which of these options is preferable for univariate data (e.g., Rogan et al. 1979); however, given multiple dependent measures, the multivariate approach is superior as a method of protecting against the experiment-wise Type I error rate (Bock, 1975).

Multivariate ANOVA for repeated measures. The multivariate approach involves the multiplication of the data (or the matrix of means) by orthogonal contrasts representing the structure of the repeated measures factor. Any orthogonal decomposition of the repeated measures effects will suffice, but for use with sequential data we recommend the Fisher-Tchebycheff orthogonal polynomials for trend (Bock, 1975). The use of a polynomial model places the multivariate ANOVA approach in the general class of polynomial growth curve models (e.g., see Guire and Kowalski, 1979; Pottloff & Roy, 1964).

We consider first the univariate case for a *CS* design. The linear model is more complex (see Bock, 1979; Timm, 1975); it essentially reduces to the following equation for the matrix of observed means:

$$\underset{\sim}{\bar{Y}} = \underset{\sim}{K}\underset{\sim}{\Theta}\underset{\sim}{P} \quad (7)$$

(Bock, 1979; Finn, 1969), where $\underset{\sim}{K}$ is a full rank basis matrix of effect contrasts for between-subjects factors, $\underset{\sim}{P}$ is a J-1 order matrix of orthogonal polynomials for trend over age, and $\underset{\sim}{\Theta}$ is the matrix of effects to be estimated. There are several methods of estimating the effects in $\underset{\sim}{\Theta}$. The simplest is to select directly a basis matrix $\underset{\sim}{K}$ of orthogonal contrasts, representing the effects of cohort, and to use the J-1 matrix of orthogonal polynomials, $\underset{\sim}{P}$. As shown by Finn (1969) and McCall and Appelbaum (1973), the procedure is then to create explicitly a new matrix of observations by postmultiplying the matrix of original observations by the orthonormalized transform of $\underset{\sim}{P}$, creating a new vector of variables, say Z, of order J. This vector is then partitioned into two exclusive parts. Since the first column of $\underset{\sim}{P}$ is a column of ones (for the grand mean), the leading element of $\underset{\sim}{Z}$ is the weighted average of all J repeated measures. An analysis of the between-subjects factorial using z_1 as a dependent measure tests the main effects and interactions of the between-subjects effects. In the *CS* design, this analysis involves the K-1 effects for cohort represented in $\underset{\sim}{K}$. The remaining J-1 variates in $\underset{\sim}{Z}$ are the repeated measures weighted by the coefficients for the corresponding polynomial terms (i.e., linear, quadratic, cubic, etc.). Under the null hypothesis of no age effects, the expected value of these weighted variates is zero; hence a test of the hypothesis that the constant terms equal zero provides a test of the main effects hypothesis for age (Guire & Kowalski, 1979; McCall & Appelbaum, 1973). The test of the effects for cohort on the transformed variates provides a test of the cohort-by-age interaction. The null hypotheses are tested by a MANOVA using the J-1 transformed variates as multiple dependent measures. The multivariate significance tests provide omnibus significance tests analogous to the omnibus tests in the mixed model ANOVA. The critical point is that testing the main effect for age is accomplished by testing the constant terms for the polynomial transforms of the original measures; usually the test of the constant term (in untransformed data) is of little interest because the magnitude of the constant is arbitrary. Examples of this type of analysis are found in Finn (1969), Finn and Mattsson (1978), and McCall and Appelbaum (1973). McCall and Appelbaum (1973) provide examples of more than one within-subjects factors as well. Schaie and Hertzog (Note 2) have used this method in the analysis of *CS* designs.

The assumptions required for the MANOVA treatment of the univariate data are much less restrictive than those of the mixed model approach—namely, that the between-subjects groups have the same general population covariance matrix $\underset{\sim}{\Sigma}$. The MANOVA approach to repeated measures analysis may be generalized to the case of P dependent variates by simply performing the transformation on the P variates simultaneously, making $\underset{\sim}{Z}$ a $J \times P$ matrix. Then the multivariate test of the hypothesis that the $(J\text{-}1) \times P$ submatrix of means of $\underset{\sim}{Z}$ is null provides an omnibus test of the hypothesis that all age effects are zero (for all P variates simultaneously).

[6] The sphericity assumption is necessary for mixed model analysis; the more frequently cited compound symmetry assumption is sufficient but not necessary—it is in fact overly restrictive (see Rogan, Keselman, & Mendoza, 1979).

Questions as to how any significant effects might be localized in some subset of polynomial terms or in some subset of the dependent measures may be handled by inspecting the univariate *F*-tests following a significant multivariate *F*, but this procedure provides no protection of Type I error rate for the multiple comparisons. A more elegant procedure is to employ step-down testing procedures (Bock, 1975; Finn, 1969), provided that the transformed variates may be ordered in such a way as to make the step-down test meaningful with respect to a priori hypotheses about localization of the significant effects. The step-down procedure consists essentially of a multivariate analysis of covariance, where the preceding *P*-1 variates are covaried on the pth variate. Ordering the variates is thus a critical part of the step-down analysis. The interested reader should consult Bock (1975, Chap. 7) for detailed discussion of the application of the step-down procedures to the multivariate repeated-measures design.

Statistical Tests of Hypotheses concerning Covariance Structures

Many developmental hypotheses are best tested by formulating statistical models regarding the structure of covariance matrices taken from longitudinal or cross-sectional sequences. As discussed in the section on statistical methods, many hypotheses regarding individual differences in developmental patterns involve examination of variances and covariances among observed measures. Furthermore the problems of measurement imprecision and the inability to manipulate human development directly have begun to force developmental psychologists to follow the lead of economists and sociologists and to consider the utility of causal modeling among latent variables by means of structural regression techniques. It is not unreasonable to expect that the next decade will see a methodological explosion in the form of increased use and appreciation of methods of modeling covariance structures by developmental psychologists, particularly those concerned with life-span developmental phenomena.

Much of the credit for this explosion, if it does indeed occur, will be given to the Swedish statistician Jöreskog, whose contributions to the methods of covariance structures analysis, especially in their application to factor analysis and structural equations models for longitudinal data, have been noteworthy. Jöreskog and co-workers, most notably Sörbom, have not only extended extant statistical models for covariance structures analysis but have also contributed an important method of statistical estimation and testing of the parameters from those covariance-structure models. We do not mean to belittle the contribution of many other scientists to the theory and methods of covariance structures analysis (see Bentler & Weeks, 1979, and Bentler, 1980, for reviews and a historical perspective on this topic). Nevertheless Jöreskog's contributions, as exemplified in a generation of computer programs designed for covariance structures analysis, seem of primary importance.

The LISREL Model. Jöreskog and Sörbom (1978) have developed a highly general model for structural equations: LISREL (LInear Structural RELations). The LISREL model consists of a system of linear structural regression equations describing the relationships among sets of observed and unobserved variables. The use of structural equations systems has been advocated as a method of assessing the putative cause-and-effect relations among correlated observed variables when experimental manipulation to achieve causal inference is not possible (see Duncan, 1975; Heise, 1975). Wright is generally credited with the development of this approach (e.g., Wright, 1954), initially termed *path analysis,* and his methods have been most frequently utilized in sociological research (Duncan, 1975). *Path analysis* is the term commonly applied when causal relations are specified among observed variables. The term *structural equations* is used when the variables in the causal system are not necessarily directly measured but may also be hypothetical construct or latent variables which may or may not be related to other observed variables (Heise, 1975).

The power of structural equations is that the usual matrix of regression coefficients among observed or latent variables is not used as an indicator of direct influences among constructs; instead, the investigator is required to specify a model regarding a causal sequence among variables. In general this will imply direct effects of some variables upon others, indirect effects for some variables upon others—implying that the usual regression coefficients reflect association through an intervening variable, and no direct or indirect effects of some variables on others—thus implying the correlation between variables to be spurious in a causal sense (see Duncan, 1975; Heise, 1975; Land, 1968). Structural equations models are particularly useful for longitudinal analyses (Duncan, 1975; Jöreskog & Sörbom, 1977; Rogosa, 1979) when certain causal sequences are known to be required by the time-structuring of the data and the causal axiom, "if *a* precedes *b*, *b* cannot cause *a*." Jöreskog and Sörbom's LISREL program is a particularly powerful method for specifying and estimating structural equations models. It represents in essence a union of restricted maximum likelihood factor analysis with multivariate structural regression equations (Jöreskog, 1973; Jöreskog & Sörbom, 1978). The procedure estimates the unknown parameters in a set of linear equations regressing endogenous dependent latent variables (the variances of which are accounted for by the model) upon exogenous independent latent variables (the variances of which lie outside the prediction of the model). Relationships may also be specified among endogenous latent variables. The latent variables are estimated through the use of maximum likelihood factor analysis. The model allows for errors in the structural regression equations (regression residuals) and errors in the regressions of latent variables on observed variables (errors of measurement). Provided that a given model is identified—that is, it is a model specifying a unique solution for all parameters—given a set of observed variables, the LISREL program will estimate all unknown regression coefficients, covariance matrices among latent variables, the residual covariance matrix, and the measurement error covariance matrices.

The LISREL model consists of two parts, the measurement model and the structural equations model. The measurement model specifies how the latent variables

(factors) are measured in terms of the observed variables; it is the factor analysis model. The structural equations model specifies the "causal" relationships among the latent variables; it is the regression model. Space limitation precludes a detailed specification of the LISREL equations (see Jöreskog & Sörbom, 1978).

There are three types of parameters in LISREL: (1) *fixed parameters* have values which are fixed in advance, (2) *free parameters* are unknowns which are to be estimated, and (3) *constrained parameters* are two or more unknowns which are constrained to have the same value. LISREL is thus a restricted model, for it is necessary to restrict (i.e., fix or constrain) a sufficient number of parameters in advance in order to uniquely identify all the freely estimated parameters. A model is identified if it produces a unique population covariance matrix $\underset{\sim}{\Sigma}$—that is, there is no arbitrary linear transformation of the LISREL parameters which produce the same $\underset{\sim}{\Sigma}$. A necessary, but not sufficient, condition for identification is that the number of unknowns in the linear equations are equal to or less than the number of observed variances and covariances. The other conditions for identification depend upon the model that is specified; treatment of the identification problem in structural equations may be found in Jöreskog (1979), Jöreskog & Sörbom (1977), Werts, Jöreskog, and Linn (1973), and Wiley (1973).

The estimation of the unknown parameters in LISREL is accomplished by maximum likelihood methods. The maximum likelihood solution is obtained by an interactive algorithm which uses the first and second derivatives of the fitting function, F (with respect to the parameter matrices) to find estimates which simultaneously minimize F. Details may be found in Jöreskog (1973). One of the chief advantages of LISREL is that the goodness of fit of the model to the sample data may be assessed by the value of the fitting function, F, at its minimum. Given the large sample assumption, F may be multiplied by the sample size to obtain a value that is asymptotically distributed as χ^2 with degrees of freedom equal to the number of elements in S, minus the number of unknown parameters fitted in the model. In exploratory situations, where the true model is unknown and several alternatives are to be compared, the improvement in fit between two models may be assessed by computing the difference in χ^2 between them, which is also asymptotically distributed as χ^2 with degrees of freedom equal to the difference in degrees of freedom between the models. This procedure is only viable if the models are nested—that is, if the parameter specification of one model is the same as the parameter specification of the other, excepting some additional free parameters. There are two important qualifications to this procedure. First, the sampling distribution of F under repeated model testings on the same data is unknown, and thus the significance tests for χ^2 have unknown Type I error rate. Repeated model modifications may be capitalizing on chance fluctuations in sample data to an unacceptable degree, and any final model should be confined (validated) in an independent sample, whenever possible. Second, the χ^2 test is highly dependent upon sample size and is sensitive to departures from multivariate normality in the data. Hence it is possible to obtain a large and significant χ^2 value when the model fits relatively well by other standards and is, in fact, an acceptable model. Absolute χ^2 should not be taken as the only or the ultimate criterion for model acceptance (Jöreskog, 1971).

The most recent version of LISREL (LISREL IV) has been extended to the simultaneous analysis of multiple groups. The chief advantage of the simultaneous analysis in multiple groups is that parameters may be fixed or constrained to equality across groups. The ability to constrain parameters across groups is particularly useful, for an investigator may then systematically test hypotheses about the equivalence of unknown parameters across groups of subjects. The basic procedure is to estimate the same model in all groups with parameters of interest constrained to equality between groups, and then to estimate the same model with the parameters free to vary between groups. Then the difference in χ^2 between the two models represents a test of the null hypothesis of group equivalance in the parameters.

Analysis of Repeated Measures Designs. Covariance structures analysis of repeated measures designs involves the use of contrast coefficients to define latent variables, for which means (effects) and variances (variance components) may be estimated. LISREL may be used as a general model to analyze repeated measures designs by specifying the model as outlined by Bock and Bargmann (1966), Jöreskog (1974), Scheifley and Schmidt (1978), and Wiley, Schmidt, and Bramble (1973). The basic procedure is to use a matrix of contrasts to define the latent variables in the measurement equations, then the variance of the latent variables are the variance components associated with the effects (see Jöreskog, 1974).

In order to analyze longitudinal sequences, multiple cohort groups must be introduced into the model. There are several ways in which this type of analysis may be performed. Jöreskog (1979) describes the analysis of growth-curve models in multiple groups by means of structural analysis, but by using a model other than LISREL. The model leaves the covariance matrix Σ unrestricted and models only the means in terms of polynomial constraints. An alternative method of estimating the growth curve model for multiple cohorts involves using the simultaneous multiple groups option in LISREL. With this approach the variance components model described above is formed by using polynomial contrasts. The variance components are estimated simultaneously across cohort groups. The null hypothesis of no cohort effects may be tested by fitting a model constraining the variance components to between-cohort equality versus a model which leaves them free to vary.

The covariance structures approach allows greater ranges of model specification than does, say, the MANOVA approach. It also does not require the assumption of homogenous covariance matrices over cohorts. Another advantage of the covariance structures approach is that the simultaneous estimation procedure eliminates the need for concern about ordering of effects to be tested—the variance components will be invariant with respect to ordering of the latent variables in the equations.

Horn and McArdle (1980) present a method of modeling additive effects models for repeated designs using structural analysis of the moment matrix. The method is similar to Jöreskog (1979) in that it is left unrestricted, only the mean vector is structured. Horn and McArdle

(1980) use equality constraints to specify the *A, C, T* effects over the different groups and occasions.

Restricted Factor Analysis. One of the more important LISREL applications for developmental research is its use for estimating a restricted factor analysis model. There are two major factor analysis applications that may be of interest to developmental psychologists: (1) analysis of measurement properties (reliability) over different age groups and (2) longitudinal factors analysis.

Measurement properties. The use of covariance structures analysis for estimation of the psychometric properties of tests has been detailed by Jöreskog (1974) and by Rock, Werts, and Flaugher (1978). The application is based upon the fact that different models for psychometric properties may be expressed in terms of a factor analysis model (Lord & Novick, 1968). The measures are said to be congeneric if a single factor accounts for the common variance of the variables. If the factor loadings may be constrained to be equal, the true score variances of the tests are equal, and the measures are said to be tau-equivalent. If both the factor loadings and the unique variances may be constrained equal, then the measures are said to be parallel forms, because both true score variances and error variances are equal (Jöreskog, 1974).

Tests of the psychometric properties of tests across different age or cohort groups is an important procedure if there is reason to suspect that the tests may have different measurement properties in the groups. Quantitative comparisons of mean differences in scores have little meaning if the tests have fundamentally different measurement properties. A minimal requirement is that the tests be congeneric measures of the hypothesized construct in all groups with the same units of measurement (scale). As shown by Rock et al. (1978), the tenability of the equivalent scales hypothesis rests upon the plausibility of a model constraining the factor loadings to equivalence between the groups. Provided that the equivalent scales hypothesis is not rejected, the groups may be compared for quantitative differences in true score means. Rock et al. provide a detailed description of how hypotheses concerning the psychometric properties of tests may be estimated using a restricted factor analysis model.

Simultaneous factor analysis in multiple groups. A similar application of restricted factor analysis involves testing the hypothesis of equivalent factor structures in multiple age groups. Hypotheses of age changes in the factor structure of intelligence and personality have been advanced by many (see Reinert, 1970, for a review). The most common hypothesis is the age-differentiation-dedifferentiation hypothesis, which states that intellectual structure is initially nearly unidimensional, differentiates during childhood and adolescence to a more complex multidimensional structure, and then dedifferentiates into a less complex structure in old age. Factor analytic studies have been cited as evidence for and against differentiation in childhood and dedifferentiation in adulthood; much of the contradictory evidence may be attributed to differing methods of exploratory factor analysis (Reinert, 1970).

A paper by Meredith (1964) bears directly upon the issue of appropriate criteria for assessing group differences in factor structure. Specifically Meredith used Lawley's selection theorem to show that, if a factor analysis model holds for a given population, then selection of subgroups from the population should still yield an invariant factor pattern matrix of raw score regressions of observed variables on factors. However, the covariance matrices of observed variables, unique components, and factors would not generally be equivalent across groups. Meredith's paper is of critical importance with regard to the hypothesis of structural change with age, for it suggests that (1) age differences in standardized factor loadings or in factor covariance matrices would be expected by age selection alone—and therefore cannot be taken as evidence that the age groups derive from separate populations in which different factor analysis models hold and (2) only group differences in the raw score factor pattern matrix constitute evidence of qualitative age differences in factor structure (Mulaik, 1972).

The LISREL model represents the ideal method of comparing groups in factor structure, since the analysis enables tests of between-groups equivalence in different parameters using equality constraints between groups and since the simultaneous analysis uses covariance matrices and not correlation matrices as input data. Examples of multiple group factor analysis of this type may be found in Jöreskog (1971), McGaw and Jöreskog (1971), and Bechtoldt (1974). Recent studies using these methods to analyze age differences in factor structure include Cunningham (1980), Horn and McArdle (1980), and Hertzog and Schaie (Note 3).

Longitudinal factor analysis. One of the major benefits of restricted factor analysis models is their application to longitudinal factor analysis. Given a matrix of multiple measures at several occasions, usual exploratory factor analysis procedures will be dominated by the high covariances among replicated measures across occasions, and will tend to recover what may be termed "test-specific" factors, one for each measure. Although such a model has some interesting properties, it does not represent the optimal model for assessing changes in factor structure over occasions. Several authors (e.g., Bentler, 1980; Corballis, 1973; Corballis & Traub, 1970) have suggested alternative longitudinal factor analysis models which better preserve the longitudinal nature of the data. The most general longitudinal factor analysis model seems to be the one specified by Jöreskog and Sörbom (1977). Their model is similar to Corballis's (1973) model except that Corballis's model is more restrictive in requiring standardized variables and factors. The general approach is to specify an occasion-specific model—i.e., one in which a particular factor analytic model is hypothesized to account for the within-occasion covariance matrix, and with the same factors replicated over occasions.

The invariance of the solution across occasions can be assessed by comparing (1) the invariance in the relationship between observed variables and factors across occasions and (2) the stability of factors across occasions. The hypothesis of cross-occasion invariance in factor loadings applies only to the maximum likelihood estimators for the unstandardized factor loadings. Standardized load-

ings will not be invariant unless the factors have the same variance across occasions. This is one of the limitations of the Corballis-Traub type longitudinal factor analysis.

The stability of individual differences across occasions is indicated in the elements in the factor covariance matrix. Differences in the factor variances would indicate that the variability of individuals around the factor mean differed between occasions. The magnitude of the factor covariances would indicate the extent of consistence in individuals' relative ranking about the factor mean; when this matrix is postscaled to a correlation matrix, then the factor correlations should approach unity as individuals approach exact maintenance of position relative to the factor mean. Thus the occasion-specific model supplies the parameters which were indicated earlier as being crucial for hypotheses regarding changes in IEV and of IEV in IAC but in terms of the latent variables. The fact that these parameters are estimated from latent variables, thereby eliminating contaminating influences of measurement error, is an important and useful property of the occasion-specific longitudinal factor analysis model. Hertzog and Schaie (Note 3), Jöreskog and Sörbom (Note 4), and Olsson and Bergman (1977) provide examples of this approach.

An important additional feature of this longitudinal factor analysis model is that it can include covariances among the residual elements. Several authors (e.g., Corballis, 1973; Sörbom, 1975) have suggested that local independence of the residuals for replicated variables is unlikely; the regression residual (or unique components) from the factor analysis model will contain both random error and reliable variance, albeit variance specific to the measure of "true" variance of a trait not common to all indicators of a factor. Failure to include autocorrelated residuals in a model where they hold in the population will perturb all other parameter estimates.

Jöreskog and Sörbom's longitudinal factor analysis model serves as the measurement model for a structural equations system designed to estimate the causal influences among latent constructs. Jöreskog and Sörbom (1977) discuss this LISREL model in detail. It has two important features: (1) an autoregressive model among the longitudinal latent variables and (2) the introduction of exogenous and endogenous predictor variables (e.g., measures of *SES,* health status). The first-order autoregressive model states that between-subjects variation about the factor mean at a given occasion $t + 1$ is predicted only by between-subjects variation about the mean of a latent variable, η, at the previous occasion, t:

$$\eta_{t+1} = \beta_{t+1}\,\eta_t + \zeta_{t+1} \qquad (8)$$

This model is consistent with a simplex (Jöreskog, 1974) pattern in the correlations among η's; correlations decrease monotonically as one moves away from the diagonal (correlations are highest for adjacent η's). This type of model has found extensive use in time-series modeling (see Frederiksen & Rotondo, 1979).

The application of multiple occasion structural models to longitudinal research is discussed extensively by Kenny (1979) and Rogosa (1979). One of the more important applications of modeling structural regressions in multiple occasion models is the cross-lagged regression model. In this model the prediction system for replicated latent variables conforms to the first-order autoregressive pattern described previously, but cross-lagged regressions are permitted between nonreplicate latent variables at adjacent occasions. The use of cross-lagged regressions is intended to isolate a causal sequence by determining which latent variable provides the greatest prediction of subsequent nonreplicate latent variables. The logic is the same as in cross-lagged correlation analysis, except that the structural regression approach has several major advantages: (1) It disattenuates the relationships among latent variables of measurement error in the observed variables; (2) it allows for simultaneous estimation of a system of lagged regressions of any order of occasions; and (3) it does not force a standardized solution, thereby preventing any cross-occasion changes in the variance of the latent variables from affecting the magnitude of prediction, as reflected in the unstandardized structural regression coefficients (Rogosa, 1979). These properties of structural regression systems make the cross-lagged regression system a better general model for studying causal influences in longitudinal data, although simple cross-lagged correlations may provide a useful "quick and dirty" test of whether the relationships are of sufficient magnitude (and of the proper form) to justify the added time and expense of structural regression analysis.

LISREL models with means. Jöreskog and Sörbom (1980) have described the introduction of means into the general LISREL model. LISREL may now be used to estimate means of latent variables, which is extremely useful for developmental analysis (Jöreskog & Sörbom, Note 4).

REFERENCE NOTES

1. Boik, Robert J. Interactions, partial interactions, and interaction contrasts in the analysis of variance. Unpublished manuscript.

2. Schaie, K. W., & Hertzog, C. Estimates of ontogenetic change in intelligence during adulthood: A fourteen-year cohort-sequential study. Unpublished manuscript.

3. Hertzog, C., & Schaie, K. W. *Aging and structural invariance in intelligence.* Paper presented at the 87th annual convention of the American Psychological Association, New York, September 1979.

4. Jöreskog, K. G., & Sörbom, D. *Simultaneous analysis of longitudinal data from several cohorts.* Paper presented at the SSRC Conference, *Analyzing Longitudinal data for Age, Period, and Cohort Effects.* Snowmass, Colorado, June 1979.

REFERENCES

ADAM, J. Statistical bias in cross-sequential studies of aging. *Experimental Aging Research,* 1977, *3,* 325–333.

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

APPELBAUM, M. I., & CRAMER, E. M. Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin,* 1974, *81,* 335–343.

BAER, D. M. An age-irrelevant concept of development. *Merrill-Palmer Quarterly,* 1970, *16,* 238–246.

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

BALTES, P. B., CORNELIUS, S. W., & NESSELROADE, J. R. Cohort effects in developmental psychology. Pp. 61–87 in J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

BALTES, P. B., & NESSELROADE, J. R. Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology,* 1970, *2,* 163–168.

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

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

BALTES, P. B., NESSELROADE, J. R., SCHAIE, K. W., & LABOUVIE, E. W. On the dilemma of regression effects in examining ability-level related differentials in ontogenetic patterns of intelligence. *Developmental Psychology,* 1972, *6,* 78–84.

BALTES, P. B., REESE, H. W., & NESSELROADE, J. R. *Life-span developmental psychology: Introduction to research methods.* Monterey, Calif.: Brooks/Cole, 1977.

BALTES, P. B., & WILLIS, S. L. Toward psychological theories of aging and development. Pp. 128–154 in J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging.* New York: Van Nostrand Reinhold, 1977.

BALTES, P. B., & WILLIS, S. L. The critical importance of appropriate methodology in the study of aging: The sample case of psychometric intelligence. In F. Hoffmeister, C. Mueller, & H. P. Krause (Eds.), *Brain function in old age.* Heidelberg, Germany: Springer-Verlag, 1979.

BANDURA, A. *Social learning and theory.* New York: General Learning Press, 1971.

BARTON, E. M., PLEMONS, J. K., WILLIS, S. L., & BALTES, P. B. Recent findings on adult and gerontological intelligence: Changing a stereotype of decline. *American Behavioral Scientist,* 1975, *19,* 224–236.

BECHTOLDT, H. P. A confirmatory analysis of the factor stability hypothesis. *Psychometrica,* 1974, *39,* 319–326.

BENTLER, P. M. Assessment of developmental factor change at the individual and group level. Pp. 145–174 in J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues.* New York: Academic Press, 1973.

BENTLER, P. M. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology,* 1980, *31,* 332–456.

BENTLER, P. M., & WEEKS, D. G. Interrelations among models for the analysis of moment structures. *Multivariate Behavior Research,* 1979, *14,* 169–185.

BIRKHILL, W. R., & SCHAIE, K. W. The effect of differential reinforcement of cautiousness in the intellectual performance of the elderly. *Journal of Gerontology,* 1975, *30,* 578–583.

BOCK, R. D. *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill, 1975.

BOCK, R. D. Univariate and multivariate analysis of variance of time-structured data. Pp. 199–231 in J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

BOCK, R. D., & BARGMANN, R. E. Analysis of covariance structures. *Psychometrica,* 1966, *31,* 507–534.

BOTWINICK, J., & ARENBERG, D. Disparate time-spans in sequential studies of aging. *Experimental Aging Research,* 1976, *2,* 55–66.

BUSS, A. R. An extension of developmental models that separate ontogenetic change and cohort differences. *Psychological Bulletin,* 1973, *80,* 466–479.

BUSS, A. R. A general developmental model for interindividual differences and intraindividual changes. *Developmental Psychology,* 1974, *10,* 70–78.

BUSS, A. R. Methodological issues in life-span developmental psychology from a dialectical perspective. *Journal of Aging and Human Development,* 1979–1980, *10,* 121–163.

CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Skokie, Ill.: Rand McNally, 1967.

COHEN, J. *Statistical power analysis for the behavioral sciences,* rev. ed. New York: Academic Press, 1977.

COHEN, J., & COHEN, P. *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

COOK, T. C., & CAMPBELL, D. T. The design and conduct of quasi-experiments and true experiments in field setting. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational research.* Skokie, Ill.: Rand McNally, 1975.

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

CORBALLIS, M. C., & TRAUB, R. E. Longitudinal factor analysis. *Psychometricka,* 1970, *35.* 79–93.

CRAMER, E. M., & APPELBAUM, M. I. Nonorthogonal analysis of variance-once again. *Psychological Bulletin,* 1980, *87,* 51–57.

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

DEVRIES, H. A. *Vigor regained.* Englewood Cliffs, N.J.: Prentice-Hall, 1974.

DONALDSON, G. *On the formulation, estimation, and testing of a developmental model of human abilities specifying age, cohort, and time parameters.* Unpublished doctoral dissertation, University of Denver, 1979.

DUNCAN, O. D. *Introduction to structural equations models.* New York: Academic Press, 1975.

ELDER, G. H., JR. *Children of the great depression.* Chicago: University of Chicago Press, 1974.

FINN, J. D. Multivariate analysis of repeated measures data. *Multivariate Behavior Research,* 1969, *4,* 391–413.

FINN, J. D., & MATTSSON, I. *Multivariate analysis in educational research.* Chicago: National Educational Resources, 1978.

FREDERIKSON, C. H., & ROTONDO, J. A. Time-series models and the study of longitudinal change. Pp. 111–153 in J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

FURBY, L. Interpreting regression toward the mean in developmental research. *Developmental Psychology,* 1973, *8,* 172–179.

GLENN, N. D. Cohort analysts' futile quest: Statistical attempts to separate age, cohort, and period effects. *American Sociological Review,* 1976, *41,* 900–904.

GOLDBERGER, A., & DUNCAN, O. D. (Eds.). *Structural equations models in the social sciences.* New York: Seminar Press, 1973.

GOULET, L. R., HAY, C. M., & BARCLAY, C. R. Sequential analysis and developmental research methods: Descriptions of cyclical phenomena. *Psychological Bulletin,* 1974, *81,* 517–521.

GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika,* 1959, *24,* 95–112.

GRIBBIN, K., & SCHAIE, K. W. Monetary incentive, age, and cognition. *Experimental Aging Research,* 1976, *2,* 461–468.

GRIBBIN, K., & SCHAIE, K. W. *The aging of tests: A methodological problem of longitudinal studies.* Paper presented at the 30th annual meeting of the Gerontological Society, San Francisco, 1977.

GRIBBIN, K., & SCHAIE, K. W. *Selective attrition in longitudinal studies: A cohort-sequential approach.* Paper presented at the International Congress of Gerontology, Tokyo, 1978.

GRIBBIN, K., STONE, V., & SCHAIE, K. W. *Ability differences between established and redefined populations in sequential studies.* Paper presented at the annual meeting of the American Psychological Association, Washington, D.C., 1976.

GUIRE, K. E., & KOWALSKI, C. J. Mathematical description and representation of developmental change functions on the intra- and interindividual levels. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

HEISE, D. R. *Causal analysis.* New York: John Wiley, 1975.

HERR, D. G., & GAEBELEIN, J. Nonorthogonal two-way analysis of variance. *Psychological Bulletin,* 1978, *85,* 207–216.

HERTZOG, C. , SCHAIE, K. W., & GRIBBIN, K. Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology,* 1978, *33,* 872–883.

HINDLEY, C. B. The place of longitudinal methods in the study of development. Pp. 23–50 in F. J. Moenks, W. W. Hartup, & J. deWit (Eds.), *Determinants of behavioral development.* New York: Academic Press, 1972.

HOFLAND, B. F., WILLIS, S. L., & BALTES, P. B. Fluid intelligence performance in the elderly: Retesting and intraindividual variability. *Journal of Educational Psychology,* 1981, *73,* in press.

HORN, J. L., & DONALDSON, G. On the myth of intellectual decline in adulthood. *American Psychologist,* 1976, *31,* 701–719.

HORN, J. L., & MCARDLE, J. J. Perspectives on mathematical/sta-

tistical model building (MASMOB) in research on aging. Pp. 503–541 in L. F. Poon (Ed.), *Aging in the 1980's: Selected contemporary issues in the psychology of aging.* Washington, D.C.: American Psychological Association, 1980.

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

JÖRESKOG, K. G. A general method for estimating a linear structural-equations system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equations models in the social sciences.* New York: Seminar Press, 1973.

JÖRESKOG, K. G. Analyzing psychological data by structural analysis of covariance matrices. Pp. 1–56 D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco: W. H. Freeman & Company Publishers, 1974.

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

JÖRESKOG, K. G., & SÖRBOM, D. Statistical models and methods for analyses of longitudinal data. Pp. 285–325 in D. S. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models.* Amsterdam: North Holland Publishing, 1977.

JÖRESKOG, K. G., & SÖRBOM, D. *LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood.* Chicago: National Educational Resources, 1978.

JÖRESKOG, K. G., & SÖRBOM, D. *Advances in factor analysis and structural equations models.* Cambridge, Mass.: Abt Associates, 1979.

JÖRESKOG, K. G., & SÖRBOM, D. *LISREL IV appendices.* Chicago: National Educational Resources, 1980.

KENNY, D. A. *Correlation and causality.* New York: John Wiley, 1979.

KERLINGER, F. N., & PEDHAZUR, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

KESSEN, W. Research design in the study of developmental problems. In P. H. Mussen (Ed.), *Handbook of research methods in child development.* New York: John Wiley, 1960.

KUHLEN, R. G. Age and intelligence: The significance of cultural change in longitudinal vs. cross-sectional findings. *Vita Humana,* 1963, *6,* 113–124.

LABOUVIE, E. W. Developmental causal structures of organism-environment interactions. *Human Development,* 1974, *17,* 444–452.

LABOUVIE, E. W. Experimental sequential structures for the exploration of ontogenetic and socio-historical changes. *Human Development,* 1978, *21,* 161–169.

LABOUVIE-VIEF, G. Adult cognitive development: In search of alternative explanations. *Merrill-Palmer Quarterly,* 1977, *23,* 227–263.

LABOUVIE-VIEF, G., & CHANDLER, M. J. Cognitive development and life-span developmental theories: Idealistic versus contextual perspectives. Pp. 181–210 in P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 1). New York: Academic Press, 1978.

LAND, K. C. Principles of path analysis. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology, 1969.* San Francisco: Jossey-Bass, 1968.

LORD, F. M., & NOVICK, M. N. *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley, 1968.

LUND, R. D. *Development and plasticity of the brain.* New York: Oxford University Press, 1978.

MCCALL, R. B. Challenges to a science of developmental psychology. *Child Development,* 1977, *48,* 333–344.

MCCALL, R. B., & APPELBAUM, M. I. Bias in the analysis of repeated measures designs: Some alternative approaches. *Child Development,* 1973, *44,* 401–415.

MCGAW, B., & JÖRESKOG, K. G. Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. *British Journal of Mathematical Psychology,* 1971, *24,* 154–168.

MASON, K. O., MASON, W. M., WINSBOROUGH, H. H., & POOLE, W. K. Some methodological issues in cohort analyses of archival data. *American Sociological Review,* 1973, *38,* 242–258.

MASON, W. M., MASON, K. O., & WINSBOROUGH, H. H. Reply to Glenn. *American Sociological Review,* 1976, *41,* 904–905.

MEREDITH, W. Notes on factorial invariance. *Psychometrika,* 1964, *29,* 177–185.

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

NESSELROADE, J. R. Issues in studying developmental change in adults from a multivariate perspective. Pp. 59–69 in J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging.* New York: Van Nostrand Reinhold, 1977.

OLSSON, V., & BERGMAN, L. R. A longitudinal factor model for studying change in ability structure. *Multivariate Behaviorial Research,* 1977, *12,* 221–241.

OVERALL, J. E., & SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 1969, *72,* 311–327.

OVERALL, J. E., SPIEGEL, D. K., & COHEN, J. Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin,* 1975, *82,* 182–186.

POTTLOFF, R. F., & ROY, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika,* 1964, *51,* 313–326.

REESE, H. W., & OVERTON, W. F. Models of development and theories of development. Pp. 116–149 in L. R. Goulet & P. B. Baltes (Eds.), *Life-span development: Research and theory.* New York: Academic Press, 1970.

REINERT, G. Comparative factor analytic studies of intelligence throughout the human life-span. Pp. 115–145 in L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory.* New York: Academic Press, 1970.

RIEGEL, K. F. *Psychology of development and history.* New York: Plenum, 1976.

RIEGEL, K. F., & RIEGEL, R. M. Development, drop and death. *Developmental Psychology,* 1972, *6,* 306–319.

RIEGEL, K. F., & ROSENWALD, G. C. (Eds.). *Structure and functions: Developmental and historical aspects.* New York: John Wiley, 1975.

RILEY, M. W., JOHNSON, W., & FONER, A. (Eds.). *Aging and society: A sociology of age stratification.* New York: Russell Sage Foundation, 1972.

ROCK, D. A., WERTS, C. E., & FLAUGHER, R. L. The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research,* 1978, *13,* 403–418.

ROGAN, J. C., KESELMAN, H. J., & MENDOZA, J. L. Analysis of repeated measurements. *British Journal of Mathematic and Statistical Psychology,* 1979, *32,* 269–286.

ROGOSA, D. Causal models in longitudinal research: Rationale, formulation, and interpretation. Pp. 263–302 in J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

ROSENTHAL, R. , & ROSNOW, R. L. *The volunteer subject.* New York: Wiley Interscience, 1975.

ROSOW, I. What is a cohort and why? *Human Development,* 1978, *21,* 65–75.

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

SCHAIE, K. W. A reinterpretation of age-related changes in cognitive structure and functioning. Pp. 485–507 in L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Research and theory.* New York: Academic Press, 1970.

SCHAIE, K. W. Can the longitudinal method be applied to psychological studies of human development? Pp. 3–22 in F. Z. Moenks, W. W. Hartup, & J. DeWitt (Eds.), *Determinants of behavioral development.* New York: Academic Press, 1972a.

SCHAIE, K. W. Limitations on the generalizability of growth curves of intelligence. *Human Development,* 1972b, *15,* 141–152.

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

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

SCHAIE, K. W. Toward a stage theory of adult cognitive development. *Journal of Aging and Human Development,* 1977–1978, *8,* 129–138.

SCHAIE, K. W. 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. New York: Academic Press, 1979.

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

SCHAIE, K. W., & GOULET, L. R. Trait theory and verbal learning processes. Pp. 567–584 in R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory.* New York: Hemisphere/Halsted, 1977.

SCHAIE, K. W., LABOUVIE, G., & BARRETT, T. J. Selective attrition effects in a fourteen-year study of adult intelligence. *Journal of Gerontology,* 1973, *28,* 328–334.

SCHAIE, K. W., LABOUVIE, G., & BUECH, B. U. Generational and cohort-specific differences in adult cognitive functioning. *Developmental Psychology,* 1973, *9,* 151–166.

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

SCHAIE, K. W., & PARHAM, I. A. Social responsibility in adulthood: Ontogenetic and sociocultural change. *Journal of Personality and Social Psychology,* 1974, *30,* 483–492.

SCHAIE, K. W., & PARHAM, I. A. Cohort-sequential analyses of adult intellectual development. *Developmental Psychology,* 1977, *13,* 649–653,

SCHAIE, K. W., & STROTHER, C. R. A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin,* 1968, *70,* 671–680.

SCHEIDT, R. J., & SCHAIE, K. W. Taxonomy of situations for the aged: Generating situational criteria. *Journal of Gerontology,* 1978, *33,* 848–857.

SCHEIFLEY, V. M., & SCHMIDT, W. H. Analyses of repeated measures data: A simulation study. *Multivariate Behavioral Research,* 1978, *13,* 347–362.

SEARLE, S. R. *Linear models.* New York: John Wiley, 1971.

SHONTZ, F. C. Single-organism designs. In P. M. Bentler, D. J. Lettieri, & G. A. Austin (Eds.), *Data analysis, strategies and designs for substance abuse research.* Washington, D.C.: U.S. Government Printing Office, 1976.

SIEGLER, I. The terminal drop hypothesis: Fact or artifact? *Experimental Aging Research,* 1975, *1,* 169–185.

SINNOTT, J. D. Everyday thinking and Piagetian operativity in adults. *Human Development,* 1975, *18,* 430–443.

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

TIMM, N. H. *Multivariate analysis with applications in education and psychology.* Monterey, Calif.: Brooks/Cole, 1975.

TIMM, N. H., & CARLSON, J. C. Analysis of variance through full rank models. *Multivariate Behavioral Research Monographs,* 1975, No. 75–1.

URBAN, H. B. The concept of develoment from a systems perspective. Pp. 46–85 in P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 1). New York: Academic Press, 1978.

WERTS, C. E., JÖRESKOG, K. G., & LINN, R. L. Identification and estimation in path analysis with unmeasured variables. *American Journal of Sociology,* 1973, *78,* 1469–1484.

WILEY, D. E. The identification problem for structural equation models with unmeasured variables. Pp. 69–83 in A. S. Goldberger & O. D. Duncan (Eds.), *Structural equations models in the social sciences.* New York: Academic Press, 1973.

WILEY, D. E., SCHMIDT, W. H., & BRAMBLE, W. J. Studies of a class of covariance structure models. *Journal of the American Statistical Association,* 1973, *68,* 317–323.

WINER, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

WOHLWILL, J. F. *The study of behavioral development.* New York: Academic Press, 1973.

WRIGHT, S. The interpretation of multivariate systems. Pp. 11–23 in O. Kempthorne (Ed.), *Statistics and mathematics in biology.* Ames: Iowa State University Press, 1954.