Methodological Problems in Descriptive Developmental Research on Adulthood and Aging[1]

K. Warner Schaie

West Virginia University

Pre-publication copy of paper
to appear in the proceedings of the
Conference on Life Span Developmental Psychology:
Methodological Issues,
West Virginia University
Morgantown, West Virginia
1971

[1]Preparation of this paper was facilitated by a grant from the National Institute of Child Health and Human Development (HD 044776-02).

Table of Contents

Contents (continued)

Methodological Problems in Descriptive Research on Adulthood and Aging

K. Warner Schaie

West Virginia University

I. Introduction

Increased methodological sophistication on the part of many developmental researchers has in the past few years led to a searching re-examination of the developmental literature. Such a re-examination suggests the blunt conclusion that much of the reasonably sound work done on children has made little contribution to the formulation of laws of human development while most of the work on adult development has been methodologically unsound (Baltes and Goulet, 1970). The most promising work with young organisms would, therefore, presently involve laboratory explorations manipulating developmental phenomena but the stage of pre-experimental descriptive work or at best the utilization of experiments of nature must necessarily continue for the study of adult behavior.

Although many of us thought that the taxonomic stage of inquiry into adult development had reached the level where experimental work could begin at about the time of publication of Biren's handbook (1959), we have since recognized that much of the literature reported there has little to offer to our understanding of aging within individuals. This literature merely reports how individuals who differ in chronological age also differ on other dimensions at a given point in time. Those few studies which are longitudinal in nature, do not help us either, since they contain a host of methodological problems, some of which will be considered in this chapter which generally reduce their status to that of biographical accounts of highly selected subgroups over unique time periods.

In my chapter in the volume to which the present is a successor (Schaie, 1970), I have attempted to show ways in which more useful information might be obtained from the existing literature by various schemes of re-analysis in the case of cognitive behavior. The present chapter, by contrast, is directed to the researcher who is about to start a study of change in adult behavior. I will attempt to raise the kinds of issues which the knowledgeable researcher on adult development must now consider before implementing his research design. By no means, do I wish to suggest that all caveats to be mentioned can be attended to successfully in any given study, but neither can they be ignored in specifying the assumptions for a study if its results are to stand up to serious critical review.

I am going to assume in my presentation that the prospective investigator has raised a meaningful research question and has defined his independent and dependent variables. My concern here then will be with issues of research implementation which although implicit in any developmental research problem will bear special consequences when working with an adult subject population. My discussion will begin with questions of sampling and generalizability and I will recommend for reasons to be detailed later than, with certain exceptions, representative samples should generally not be used in studies of adult development. I will then plead for the utility of hypothetico-deductive models in studies of adult development and review the consequences of alternate models for research implementation.

Without repeating the detailed presentation of the general developmental model (Schaie, 1965) I will take another tack at the age differences vs. age changes issue, by placing this issue in the context of distinguishing research questions which are designed to consider within generation changes from those involving matters of species differentiation. I will try to clarify this basic issue further by distinguishing the implications for strategy of research effort involving cohort-specific changes, species-specific changes and culture-specific changes. Each of these models, as will be seen, requires alternate sets of restrictive assumptions to permit meaningful interpretation of research results. I will then examine some special issues occurring in within cohort (longitudinal) research designs, including the effects of socio-cultural change, effects of repeated measurement, and changes in factor structure over time. Similarly, between cohort (cross-sectional) designs will require consideration of comparison within and between cross-sectional studies, prospective and retrospective estimates as well as the question of whether age-appropriate or cohort-appropriate instruments need to be used to measure one's variables. Finally, I shall provide some comments on what I see as the hierarchy of priorities to be assigned to the above issues to achieve an orderly forward move out of our present methodological dilemmas.

II. Some Questions Related to Sampling and Generalizability

It has been conventional in the psychological literature which addresses itself to the mapping of the nature of individual differences to demand that samples ought to be as broadly representative as possible in order to allow wide generality of research findings. Many investigators therefore are quite apologetic in describing their more limited samples, concluding inevitably that their findings ought to be replicated on more representative samples. This stance is a very simple minded one in that it assumes that there are some valid a priori reasons why a representative sample is indeed required to solve a particular problem. Although there may be some rationale for the representativeness of the investigation in survey and demographic studies, it may be questioned whether or not a representative sampling design is always the ideal to work towards in developmental research with adult subjects. Before answering this question one way or the other we must first of all understand what kind of age changes we wish to study, i.e. whether the reference population we wish to generalize is to be the intact, representative, idealized or restricted population going through a temporal sequence. Next we need to consider whether we are interested in developmental change occurring in the organism either as a function of species contingent maturational events and constant environmental inputs or whether we want to study the effects of cumulative pathology whose prevalence increases with advancing age but whose incidence is not inevitable in any individual member of the group under investigation. We will also see that the model of aging to be utilized will help us determine which of the possible sampling strategies appears most reasonable and likely to yield data which will provide a valid test of our model.

A. What age changes are to be studied?

The student of adult behavior has at least three different options for selecting his subject population. These options arise out of the fact that attrition of samples is inevitable due to the correlation of age with mortality as well as the fact that there are only rare populations of adults which are captive in the sense that the researcher can prevent their incidental or purposive removal from further participation in his investigation.The researcher may, therefore, select his sample to be representative of the population of individuals to which he wishes to generalize at the study's inception, he may consider data only for an intact sample (i.e. those individuals who are members of the residual sample at the study's conclusion), or he may be quite selective in specifying a sample which is appropriate to the variable to be studied, but not necessarily representative with respect to any identifiable population group as such.

1. Aging in representative populations.

The definition of a representative sample is fairly straightforward in a longitudinal study. All that is implied here is that random (or stratified) sampling has occurred from a census type listing of all the members of the population cohort at a specified chronological age within the demographic limitations thought to be reasonable for a given study. Generalization here may be feasible to wider populations dependent upon our knowledge of the demographic characteristics of the sampling base. Note, however, that except for very unusual circumstances of environmental stability or experimenter control over his sample, the representative nature of the sample will hold true only for the first measurement occasion and not thereafter. This is necessarily the case, since sample attrition due to natural causes or experimental artifacts will rarely be random. Two radically different sampling strategies should, therefore, be advocated. The choice of strategies depends on whether the investigator's objective is to assess age changes within individual organisms or whether he wishes to estimate population parameters of age changes to develop general laws.

In the first instance, in the absence of total control over one's subjects, it will be necessary to conduct collateral studies on the effects of experimenta attrition (to be discussed in more detail further on) to determine in what manner the sample deviates from its original representative nature at subsequent measurement points. But also, and this caution holds equally true for samples which suffer no attrition whatsoever, it will be necessary to determine whether the residual sample at subsequent measurement points is still representative of its parent population. Developmental researcher often neglect the fact that the characteristics of the parent population change through attrition over time in a systematic manner, but not necessarily in the same manner as holds true for a given longitudinal sample drawn from that population. It is obvious then that longitudinal samples cannot be legitimately described to be representative of a given population at more than one point in time unless suitable collateral evidence is presented. In any repeated measurement study, moreover, a sample ceases to be representative of its parent population as soon as it has been tested once in the sense that it's response characteristics have of necessity been modified by the initial assessment in a manner which is not characteristic of the parent population. That is, at subsequent measurement points it would differ from any other random population sample in that the assessment tools or situation had previously been applied to the longitudinal sample but to no other

If the investigator is interested in estimating population parameters witho concern for the evaluation of individual age gradients he should then consider t independent random sampling approach to longitudinal study. Specifically, one would take a random sample from a given cohort at Time 1, and at subsequent measurement points would compare it with new random samples drawn from the same cohort. Depending on the mobility characteristics of the population base, this would require application of either a sampling with or without replacement model

It is suggested that the latter model is perhaps the only case in which use of a representative sample for the study of developmental questions can at all be logically defended.

It is possible theoretically to obtain a representative sample of the total population at a given point in time in a cross-sectional study. The sub-sample in such a study for a given age, however, cannot be representative of the total membership of the particular cohort. This can be demonstrated by inspection of actuarial tables which show that in adult samples, differential attrition occurs for successive cohorts at identical ages. This means that it would be hazardous to argue that a sample of fifty year olds tested in 1970 is as representative of people born in 1920 as the sample of fourty year olds is representative of people born in 1930. For that matter, we cannot assume either that an obtainable sample of fifty year olds in 1970 is likely to be as representative of their cohort as a similar sample of fifty year olds would have been had we assessed such a group in 1960.

In view of the fact that there are intense generation and age differences in many demographic variables it is simply misleading to apply the term representative with respect to a cross-sectional study. A sample which is representative with respect to one demographic variable will by definition have to be biased with respect to other characteristics. These matters should be obvious, but it is timely to bring them into focus, if only to educate journal editors who continue to insist on broadly representative sampling bases for studies of adult development without full awareness of the absurdity of their requirement.

To summarize, samples representative of a given cohort at birth will successively become less representative of that cohort because its attrition will not necessarily follow the laws applying to attrition in the population. A cross-sectional sample of a population at one point in time cannot have generality because rates of cumulative trauma and other factors effecting attrition will change from one generation to the next.

2. Aging in intact populations.

Having rejected the utility of representative sampling for developmental studies of adult behavior, except in the special case of the longitudinal independent random sampling design, we must now turn to the major alternative. That is, one might select a sample with respect to its stability characteristics, so that one might study only those individuals who will survive the duration of a given study. While this approach is feasible in terms of criteria such as predictors of longevity if one wishes to study behavior changes in the very old (of. Platt and Parkes, 1967), it generally suffers from the problem that we do not generally have adequate knowledge of the factors which permit predicting whether or not a given subject will remain in a study. Consequently, the study of intact populations is frequently defined either as the retrospective study of a presently available sample, or as the analysis of data restricted to the survivors of a given longitudinal sample. The generality of findings in such study is impaired again in the first case by the cautions suggested for the representative cross-sectional sample, and in the latter case requires the usual collateral studies regarding the differences between survivors and non-survivors.

If population parameters rather than individual age gradients are of interest the further option exists to treat the survivors as a random sample of their population cohort at the latter measurement point and compare their average with the mean for the total sample at the study's initiation. This approach, however, is probably a poor substitute for the independent random sampling approach advocated earlier. Most longitudinal studies, however, will simply be limited by the nature of their sample of survivors, and findings reported therefrom should be carefully identified as being restricted to similar populations.

If the experimenters' objectives are well defined it may consequently be better to select extreme samples which permit explication of certain relationshi albeit for very restricted population bases. When population parameters are to be estimated, however, it seems clear that sequential strategies permitting generalization over several generations will be required, and that stratified sampling should be considered with respect to demographic characteristics which have low correlation with the variables to be investigated, if random or near-random selection with respect to the experimental variables is to be hoped for.

3. Aging in extreme instances.

There are many variables where the experimenter is not interested in population parameters but wishes to study how various behaviors express themselves at different developmental stages under idealized or stress conditions. In these instances one might argue for the desirability rather than the disadvantage of highly biased samples. For example, if one wants to gather data on optimal intellectual function in old age for the purpose of developing retraining programs, it would seem better strategy to work with a sample of the active intelligent and intact aged to determine what the maximal parameters might be within which individuals at a given age level can perform.

The use of extreme samples is not without its own problems. In contrast to the representative sampling situation, where as heterogeneous a sample as possible is sought, the extreme sample should be as homogeneous as possible. The special pannel, or highly selected volunteer group is therefore a most useful source of subjects. Depending upon the variables to be assessed, careful screening of subjects may be required. If optimal performance levels are to be sought, selection procedures may need to include routines to discourage all but the most highly motivated volunteers. Conversely, if one is interested in defining minimum levels of functioning compatible with survival, institutional populations may be ideal. An explicit example of a study employing the strategy of selecting an extreme sample has been reported by Schaie and Strother (1968c).

Although the question of generality is not at issue with extreme samples, it is still necessary to know the relation of the sample to its parent population. One may need to be concerned about the issue of base rates, that is the incidence of the extreme subjects in the population, as well as the extent to which the subjects investigated are skewed from the population average. The latter concern is of particular importance when subsequent parametric studies on the entire range of the behavior of interest is contemplated.

Even in extreme samples, the experimenter will face the question whether he should randomly sample from the class of extreme subjects or whether he should narrowly define his class so as to permit exhaustive study of a finite population. If he uses the random sampling approach he will again face the dilemma whether to base his findings on data restricted to those subjects available at all measurement points or whether he should compare the means of his subjects at all points as being representative samples of the parent population. No such dilemma exists if the extreme sample is treated as a finite population. In that case, measures of change must be restricted to the survivors only, since it would be absurd to include data on individuals who are no longer members of the changing finite population. Measures of population status, however, would by similar reasoning be based on the total population at each measurement point.

B. Developmental Change vs. Development of Pathology

Most developmentalists interested in maturity and old age implicitly accept a model of aging which entails cumulative traumatic experience and loss of function due to pathologic occurrences which may or may not be causally related to the species-determined biological clock or the mere passage of time. Nevertheless, most sampling plans either ignore such a model or explicitly specify selection criteria which are at best arbitrary conventions. But, if we are concerned about the generalizability of our research findings, we must also consider the impact of selection decisions which determine how pathology is to be treated.

A common procedure is to specify that no subjects will be included who suffer from physical disabilities or who are currently under psychiatric care. In view of the fact that the prevalence of physical disability and psychiatric disorder increases systematically with age, such selection would ensure differential representation of successive cohorts in cross-sectional studies and would lead to systematic experimental attrition in longitudinal inquiries. That is, if it were true that decremental developmental changes are exacerbated by physical or emotional trauma, then the changed reported in "normal" populations defined as above would yield serious underestimates of expected developmental change in samples assessed at age levels with low prevalence of pathology.

A similar objective, that of controlling for the effects of social patholog is often approached by excluding institutionalized subjects from a study. Again the representativeness of the sample will be impaired since the frequency of institutionalized members of successive cohorts will show a substantial correlation with chronological age if only because of the statutory provisions for handling disabled and aged individuals. Whether or not such exclusion is defensible will again depend upon the experimenter's principal objective. That is, if population parameters are to be estimated, then the institutionalized population ought to be sampled as well. If not, collateral studies may be in order to determine in what manner institutionalized and non-institutionalized populations do differ with respect to the variable of interest. This issue becomes important when comparisons of older and younger subjects are to be made, and differences are found which parsimoniously cannot be attributed either to age generation or environmental treatment effects, but which will be due simply to the comparison of two demographically non-equivalent samples.

It may, of course, be argued that certain types of physical impairment or institutional restraints may result in the artifactual impairment of psychological functions. The a priori assumption, however, that this is indeed the case may be based more on the investigator's hunches than upon empirical data. Indeed, much of the more respectable experimental aging research has been conducted with institutionalized and physically inpaired samples. Failure to replicate findings has usually occurred because of specific design flaws rather than the usual characteristics of the sample employed.

In view of the fact that the frequency of pathology increases with age, and that cumulative effects of pathology may be noted in all aging individuals, it would seem futule to trace "pathology free" developmental processes, except in cases where it can be demonstrated that occurrence of a given pathology does not correlate with chronological age. In the latter instance, experimental or statistical control of the degree of such pathology is, of course, a must. The emphasis of the researcher concerned about the impact of pathology as a confound in aging study ought therefore be directed to the study of that prevalence of pathology at different ages in relation to the variables he is interested in.

C. Models of Aging and Resulting Research Strategies

If generalizations are to be drawn from studies of adult development then models of aging must be implicit in the investigator's design and subsequent sampling plan. It would be better if the investigator were to explicate his model while he is still in the design phase rather than attempt to perform a pos hoc analysis to determine what kind of model his research design had expressed intuitively. Two different, but equally important, issues must be met in the design of any developmental study. First, the investigator must specify the dimension of development which is to be investigated. If he is interested in generalizing over maturational events he will most likely design a project which investigates age changes within generations. If he is interested in describing the performance of organisms at different maturational levels at one historic point in time, or to prepare data basic to social policy decisions to be implemented immediately, he will then be concerned with age differences. On the other hand, if he is concerned with the stability of behavioral phenomena at given levels of development, or with the effect of treatment variables upon developmental phenomena which are independent of maturational events, he will have to consider a time-lag approach. The implications of this major question will be discussed in more detail further on.

The second basic question which should be explicated in any developmental research plan is the nature of the assumed age function. Although it might be attractive to plead ignorance and make no assumption whatsoever, the result of such decision for good research design is clearly unfortunate. Such stance implicitly leads to designs which seek to refute the null hypothesis! A much better approach would seem to be the specification of an experimental hypothesis or alternative hypotheses, with subsequent test whether or not observed data fit the specified function.

Three different kinds of models for age functions appear to underlie most empirical investigations. The first assumes that once maturity is reached, adult behavior remains stable. The second model assumes that aging in adults is accompained by decrement phenomena, such that in an irreversable manner, the old are disadvantaged as compared to the young. The third model also assumes that there is decrement, but qualifies the "irreversability" assumption of the second model, by introducing the additional condition of cumulative environmental compensation. Each of these models allows further elaboration, since "satbility" may involve cycling within a range and decrement functions can have a wide variety of shapes. The question about the nature of the developmental function is equally appropriate whether one investigates age changes, age differences or time-lags. In the latter two instances, it would, however, be more precise to speak of generation gradients or time of measurement (historical time) gradients, rather than age functions.

1. Stability of adult behavior.

Stability of adult behavior can occur for a variety of functions which may be of interest to the psychologist. I have previously pointed out (Schaie, 1970) that the "stability" model is quite reasonable for measures of the chrystallized abilities, where it is assumed that the organism at maturity has acquired all the information available in the environment with the respect to the measurement variable, and where there is no reason to suspect that the ratio of new information input to old information obsolescence will seriously deviate from unity. If the information gain/loss ratio (GLR) is above unity slight within generation increment is to be expected while a GLR below unity will create the appearance of a slight decrement function. Secular variations of the GLR ratio, moreover, will lead to age functions which appear cyclical in nature.

It should be noted that the "stability" model is suitable also for those few behaviors which are primarily biologically mediated, but whose biological base remains stable from maturity to death. Reversible effects caused by temporary cell or tissue loss and their regeneration would again result in cyclical within generation functions.

Note further that the "stability" model implicitly assumes that the normal state of the organism is stability from maturity to death and that the presence of irreversible changes are incompatible with such a model. To adopt this model, therefore, the investigator must either specify that performance on the variables of interest is uncorrelated with pathology, or he must be prepared to study pathology-free, and consequently unrepresentative samples with respect to chronological age.

It should not be thought that age functions suggested by the stability model are identical or immutable. We have already noted that such functions can imply the maintenance of a constant level from maturity on or the cycling about an optimal level as a function internal and external events. The functio can show changes also in the age at which the asymptotic level is achieved, in the age level at which performance begins to be correlated with pathology or where pathology-free subjects can no longer be found.

In any event, once the "stability" model is postulated, the investigator is no longer concerned with within generation age functions or with age differences, since these are specified to be zero for the age range of concern. Instead one would be concerned with the determination of generational shifts in the asymptotic level attained at maturity and differentiate such shifts from th transient cyclical events occurring as a function of time-specific inputs. The cross-sequential method (Schaie, 1965; Schaie and Strother, 1968a) would then become the experimental strategy of choice, as it permits differentiation of cohort and time of measurement effects, given the assumption of zero age differences.

2. Irreversible decrement.

If we were to consider the entire life-span, we would need to distinguish between developmental functions which are linear or curvilinear, and among the latter those which are concave and those which are J shaped. There are, of course, certain functions, which show cumulative decrement from birth on (e.g. number of cells in the central nervous system), but for most variables of interest to the psychologist it is possible to demonstrate a point of peak performance which typically (although not always) occurs in young adulthood. The "irreversible decrement" model, of necessity, assumes that performance is to be charted from that peak point to death. It also assumes, similar to the "stability" model, that the effects of pathology are uncorrelated with the variables of interest, or, conversely to the stability model, that the investigator's sample at all age levels contains a representation of pathology characteristic of the general population. The latter assumption is particularl important if it is thought that decrement is to be conceptualized as the result of cumulative pathology.

A further problem with the "irreversible decrement" model, is the need to know the peak point of performance in univariate, or the range of peak performances in multivariate studies. But such peak performance ages do not remain stable, since they may be subject to generational shift also.

Perhaps the most useful application of the "irreversible decrement" model is for those variables where performance seems significantly dominated by peripheral sensory functions and psychomotor speed. It may be argued that age functions here are likely to follow a biologically determined pattern which is little influenced by the cyclical events occurring within the life-span of a single organism, even though the function may shift as a consequence of events occurring to the species or any of its subgroups over long periods of time.

Since the irreversible decrement model specifies that age changes will occur as a function of maturational events, regardless of environmental input, it becomes necessary to differentiate the variance which is maturationally determined from that which may be generation specific, that is, level and attainment of peak performance age. The cohort-sequential method has the attributes necessary for this differentiation. However, Wohlwill (1970c) has highlighted the reasons which make this method the least practicable strategy for developmental studies, and it follows that the "irreversible decrement" model is one that can be tested only with considerable difficulty. From a psychological viewpoint moreover, it does not seem probable that environmental input will be of minimal importance.

Cycling effects here, if present, would tend to yield pseudostability functions for high gain–loss ratios and very steep decrement functions where the gain–loss ratio is low.

C. *Decrement with Compensation*

A more reasonable position than the "irreversible decrement" model may be one which assumes that maturational events require phases of growth and decline which can be approximated by quasi-linear functions for either portion, but where curvilinear alternatives exist which do not require that the peak of performance occur temporally at an early adult stage. Moreover the curvilinear patterns may be explained by environmental compensatory input, which becomes effective during the decrement, but not during the increment phase of a life-span growth gradient. In other words, during the growth phase, environmental input affects observed performance minimally, since the organism has not yet reached its capacity reached its capacity as specified by its biologically determined limits. During the decrement phase, however, biologically determined decrement may in part be compensated for by environmental input, at least as observed in complex psychological measures.

The "decrement with compensation" model may be particularly appropriate for measures of fluid intelligence and for many other psychological variables where reaction time is involved. The model is also generally appropriate for variables where age changes during the adult period must be expected, where long-term shifts over generations seem unlikely, but where environmental input should have important cycling effects. For all such variables it would seem important to differentiate variance due to maturational events from the variance due to transient environmental input which might conceal the maturational process. The strategy of choice for such differentiation would generally be the time-sequential method, the most practical of the sequential research strategies, which allows identification of age and time of measurement variance under the assumption of zero generation differences (Schaie 1965; Schaie & Strother, 1968b).

It should be noted that the last model can be applied to both intact or representative samples, since the presence or absence of age-related pathology is not critical. That is, pathology is assumed to be present due to biologically mediated decrement events but such pathology may be compensated for by environmental inputs. Consequently a case could be made here to study a variable in the fully compensated case (no observable pathology) or in the naturalistic case uncompensated pathology is sampled by a representative design).

Depending upon the slope of the underlying maturational gradient, we may either expect for this model that a moderately accelerating decrement gradient be observed, or that in the extreme case, decrement be noted only just prior to death, that is, when compensatory inputs no longer suffice to stabilize the behavior in question. The latter situation would, of course, be difficult to differentiate from the "stability" model. Indeed, a reasonable design would collect data in such a way that the stability and compensated decrement models could be examined as alternate experimental hypotheses.

III. Within Generation Changes and Species Differentiation: Should One Study Age Changes or Age Differences?

We must now backtrack to the fundamental question of what it is that the developmentalist wishes to study, with special application to the analysis of the adult phase of the human life cycle. In discussing various models of aging we have already indicated that different classes of variables may require different assumptions about their developmental functions, which in turn require different experimental strategies. Overriding these concerns, however, is the question whether the developmentalist should concern himself about changes over time within organisms or the comparison of organisms at various developmental stages at one point in time. In the past it has seemed as if the latter concern has simply been one of approximating data which could not be directly obtained for the former. I would now like to propose that the questions to be asked by either approach are, although related, quite distinct and of equal social importance. To show the relevance of this issue to the design of studies on adult development this section will summarize the formal relationships between the two approaches (the general developmental model). I shall then formalize the additional restrictions desirable for the period of adult development which follow from my earlier discussion of the relationship of age function models to sampling strategy.

A. The General Developmental Model Restricted to the Period of Adult Development

The General Developmental Model simply holds that the magnitude of a response (as studied by the developmentalist) is determined by the age of the organism, the cohort (generation) to which the organism belongs, and the point in time at which the response is measured. Although the three components of developmental change are theoretically independent, they are confounded in the sense that the index for the third component is defined by knowledge of the other two indices. Since there are three ways in which three components can combine two at a time, it follows from the general developmental model, that there must be three ways in which empirical data can be collected, each approach resulting in different although related confounds. The three ways in which developmental data can be studied are, of course, the longitudinal, cross-sectional and time lag methods (Schaie, 1965). Each approach confounds two components of developmental change. The longitudinal confounds age and environmental impact between times of measurement,the cross-sectional confounds age and generation differences, and the time-lag method confounds environmental impact between times of measurement and generation differences.

The general developmental model further calls attention to the fact that the investigator has at his disposal for any particular range of the life span a trapezoidal population, consisting of several generations, available to the investigator over a specified period of time. The investigator therefore has the option to study several cohorts over a specified age range (cohort-sequential method), several cohorts over a specified period of time (cross-sequential method) or several age levels over a specified period of time (time-sequential method). The traditional methods, given the assumption that one of the components has a zero value, by comparison can, of course, yield information on one of the three components only. The sequential methods have the advantage that given certain assumptions, they will yield information on two of the three components of developmental change. Additional assumptions may also permit the assignment of theoretical values to all three components.

Since the age function for any variable would be determined by the combination of functions for the three components of developmental change, there must be as many as 729 distinct models for any growth function having as few as three observation points, if we allow nine distinguishable gradient slopes (zero, positive, negative, concave, convex, accelerating, decelerating, positive asymptotic, negative asymptotic). All these models will need to be considered for total life-span studies. But if we define adult development, as the period beginning with the attainment of maximal maturation, then matters become conside ably simpler. We may immediately neglect all models for the age components of developmental change which specify positive, concave, convex, accelerating or possitivex asymptotic slopes, resulting in a reduction to 324 distinct models. If we further assume that for a given variable environmental input (whether shor or long range) is uni-directional, that is either good or bad, we then find ourselves limited to the manageable number of, at most, twenty-four alternative component combinations which would explain variation in adult behavior. Explicitly, the stability model would permit sixteen combinations of the cohort and time of measurement gradients (zero, positive, accelerating and positive asymptotic for each). The "irreversible" decrement model would allow four combinations (negative for age,with either zero, positive, accelerating or positive asymptotic for the cohort gradient), as would the decrement with compensation model (negative for age, zero, positive, accelerating or positive asymptotic for environmental input).

B. Some Special Cases Requiring Further Restrictions

What then are the related but conceptually distinct questions which the developmentalist can and should address himself to.[2] I have already stated that there are three kinds of questions which can be deduced from the general model. The first is concerned with cohort-specific changes, or the ontogenetic study of individual organisms. The second involves the study of species-specifi changes, or the variation in developmental functions for classes of organisms over generations. The third involves the study of culture specific changes. That is, the way in which changes in the environment will modify the response of organisms at equivalent age levels. The particular confound in which the components of developmental change must be observed to bear upon these questions will now be specified and examples of psychological questions given for each ty

1. Cohort-specific changes: The exploding universe of knowledge and the chrystallized abilities.

A basic question to which many developmentalists wish to address themselves is that of predicting change of behavior over time from one developmental stage to the next. For those variables where there is relatively low variability in the slope of the developmental gradient this question can also be phrased as that of the description of developmental change within a homogenous group of organisms. The conventional longitudinal repeated measurement study addresses itself to this question, but it can answer the question only in a historical manner for the specific cohort which is being examined, unless the absence of cohort and time of measurement effects can be demonstrated. Since these assumptions are quite unreasonable for most psychological variables in adults I have, therefore, argued that the conventional longitudinal study may be suitable for biological but not for psychological data (Schaie, 1971a). This does not mean, however, that the basic question is trivial for the psychologist. In those instances where there are maturational levels we need to know what they are, and in those cases where environmental input is all important we must know how this affects development. But while models can be built from the study of individual cases, nomothetic prediction will require adequate sampling of successive populations for any developmental study. Given such information we can then specify, what kind of age changes ought to occur for members of a specific cohort, respectively for a given individual.

The case of the chrystallized abilities can serve as an example. In general, cross-sectional data show moderate decline from an adult peak (e.g. Schaie, 1958) while longitudinal data show lack of decrement (e.g. Owens, 1966). From the general model it follows that such findings should suggest within cohort stability but between cohort differentiation, conceivably modified by cycling of the information GLR. Let us now assume that there is a marked increment in the amount of information available to successive generations, a notion consistent with the explosive increase in the universe of knowledge. If we, moreover, accept Cattell's (1965) hypothesis that the chrystallized abilities are acquired, we would then have to argue that the longitudinal approach will not provide us with information useful for prediction once we have accepted the adult stability model. As discussed earlier, it will now be necessary to apply the cross-sequential model to test whether the consequences of an exploding universe of knowledge will in fact lead to age differences in asymptotic levels for subsequent generations and furthermore whether temporal cycling of the GLR does occur. Such data, if sampled over several cohorts and times of measurement would then yield the functions which would permit cohort-specific forward prediction. A longitudinal measure would provide information on the cycling (due to T) which had occurred in the cohort on which the information was gained, but not the interaction with level of performance which is needed to predict the performance of a subsequent cohort at a given age level. The appropriate developmental confound in the latter case is, of course, the time lag.

The analysis of cohort-specific data is appropriate whenever we wish to give an historical account of the behavior of a specific generation (by means of longitudinal analysis) and when we wish to predict or make social policy provisions for the future behavior of a particular population group (by means of time-lag analysis). It does not seem appropriate for the specification of age functions thought to be general laws.

2. Species-specific changes: The case of the fluid abilities.

A quite different question is being asked when variations in age functions which are maturationally determined are to be traced over time. What we are interested in here is to determine whether or not the same growth and decrement functions will hold for successive generations. The very practical reason for asking such a question might be to determine whether (a) individuals need to receive differential treatment at different age levels at the present time and (b) whether policies once determined are likely to require modification in the future. In contrast to cohort-specific questions, we are here interested in generalizable age functions, but also how such age functions vary over time. Psychological variables to which these questions apply in the study of adult behavior are generally those where speed of performance is important. As an example take the case of the fluid abilities, where it may be argued that the matrix of development is largely set by maturational determinants, but where the level of accomplishment may vary over generations due to either genetic shifts or modifications in the environment equally effective for all members of the population under study. In this particular instance we would need to apply the cohort-sequential model in order to differentiate the nature of the age function from generational shifts occurring therein, i.e. we would want to sample over ages and generations to obtain general laws. Our specific questions with respect to the shape of the age function, however, could be answered by a single longitudinal study if we assume cohort-specific environmental impact to be zero. The question of differences between age level at the present time, however, would need to be answered by a cross-sectional study, since the confound of age and generation would be required to respond to the particular question. Species-specific data from a single generation will in this case, therefore, suffice to give information on the shape of the age function, but sequential designs will still be needed to give information on generalizable levels of the function.

3. Culture-specific changes: The case of disengagement.

A further question of interest to the developmentalist is to determine whether a given behavior thought to be age-related occurs systematically and at a specified level at a given age. We may want to know whether the observed behavior could be an artifact of a particular measurement occasion or whether the level of performance at a given age will change from one measurement occasion to another. This question, incidentally can also be broadened to apply to any kind of behavior established by operant procedures (cf. Baer, 1971), since we must know whether such behaviors can reliably be programmed at a given developmental level, or whether the original success was an artifact of non-recurring environmental circumstances. A more familiar example to workers interested in adult development, however, would be the case of disengagment. Here, we have an explicit theoretical formulation, which specified that reduction in the level of social involvement will occur as an adaptive response in old age, and that such a response has been programmed by the role expectations in our culture. Unless we further assume that such role expectations are also species-specific gives (a clear absurdity) then we would have to conclude that neither a cross-sectional nor a longitudinal study could provide valid evidence to support disengagement theory. In the one case we would simply answer the question as to the level of disengagement at different ages at one point in time, while in the other we would specify if and when disengagement occurred within a single generation. But in both instances we would have to assume a priori that disengagement as an age related phenomenon does exist. To ask the more interesting general question, we would require a time-sequential strategy where we would sample different ages at different times of measurement. Only in such a manner could we conclude whether or not disengagement was indeed a general phenomenon occuring at a specified developmental level. Otherwise we are left with the more parsimonious alternative that disengagement occurs in individuals having certain sociological attributes which are unrelated to aging, even though they occur with higher frequency in the aged (cf. section II B above).

C. Implication of Models for the Selection of Research Strategies

The plea for the selection of the appropriate research model in studies of adult development which the above examples have sought to focus has been made in the realization that past descriptive research has frequently collected the wrong data to answer the right question, or even worse collected the right data in ignorance of the questions to be asked appropriately from such data. I shall now summarize this issue by stating unequivocally that a proper research strategy and design for data collection can be developed only if (a) the specific developmental question is made explicit and (b) if the hypothesized shape of the age function is specified. The first constraint is required to know what data are to be collected and should be obvious. The second is less obvious but must be required if developmental implications are to be inferred from the data. It is argued then that descriptive developmental research must utilize a hypothetico-deductive approach, because observational data necessarily confound dimensions of developmental change and do not permit the inductive procedures permissible in the non-developmental sciences.

As a guide to the reader who may find it difficult to deduce a "code of conduct" from the series of limiting assumptions specified in this and the preceding section of this chapter, I will now summarize the data gathering strategies which I consider most appropriate for each of the descriptive questions discussed above under the assumptions necessary for each of the three major models for adult development:

1. Cohort-specific changes:

a. Are there age changes in general? This question cannot be answered by any cohort-specific strategy.

b. Are there age changes for a specific cohort (post-dictive)?

i. Stability model. We assume that there are no age changes as such, but a longitudinal study would yield information on the GLR cycling attributable to time of measurement differences.

ii. Irreversible decrement model: Since time of measurement differences are here assumed to be zero, the longitudinal method would provide the age function for the specific cohort.

iii. Decrement with compensation model: Time-sequential method to permit differentiation of change due to maturation from that due to specific temporal compensatory inputs.

c. Are there age changes for a specific cohort (predictive)?

i. Stability model: Since no maturational changes are expected we need the confound of generation and time of measurement differences to predict the future performance of a specific cohort under this assumption. This would involve time-lags and consequently the cross-sequential method is needed to sample enough past time-lags to provide realistic estimates.

ii. Irreversible decrement model: Here we wish to forecast maturational changes which may differ among generations. The cohort-sequential method is therefore required.

iii. Decrement with compensation: Performance of a specific cohort at given ages may vary due to changes in environmental input. Estimates must therefore be based on samples at various ages and times of measurement, the time-sequential method.

2. Species-specific changes:

a. Are there age changes in general?

i. Stability model: No age changes are postulated, but the cross-sectional method would give information about generation differences.

ii. Irreversible decrement model: The cohort-sequential method would be required to describe both age function and generation differences therein.

iii. Decrement with compensation model: Since compensation has been postulated as trivial with respect to species-specific changes, and generation differences as trivial to the compensation model, the longitudinal method should in this case provide an acceptable age function.

b. Do individuals at different ages require differential treatment at the present time?

i. Stability model: Different treatment would have to be based on the fact of generational differences. These would in the absence of age differences be directly detected by the cross-sectional method.

ii. Irreversible decrement model: Present differences could be due to age differences, cohort differences or both. Since they do not require differentiation to answer this question, the cross-sectional method is again appropriate.

iii. Decrement with compensation model: This model postulates the absence of generation differences and the cross-sectional method is therefore the proper estimate of age differences.

c. Will social policies with respect to age require future modification?

This question is directly concerned with the issue of generation differences. Solutions require the differentiation of age and generation change with identical designs provided for the general question o age changes in section 2a above.

3. Culture-specific changes:

The basic question here is whether or not individuals at given ages show behavior variations as a function of the point in time when the behavior is observed. This is generally speaking a question of measuring time-lag . The postulated age function model will nevertheless be of importance in planning the data collection if time-lag is to be investigated for more than a single age level.

a. Stability model: Since no age changes are postulated a simple time-lag design for each of the age levels and points in time of interest will suffice. For more than one age level this is formally equivalent to either cohort or time sequential method with intensive sampling at each age level over either cohorts or times of measurement.

b. Irreversible decrement model: Here cultural changes in the form of generation differences must be separated from age changes. The cohort-sequential method would be most appropriate.

c. Decrement with compensation model: In this instance we need to separate the effects of age from those of time of measurement differences, and the time-sequential method is the design of choice.

4. What can the descriptive developmental strategies tell us?

At the risk of being unnecessarily redundant I will end this section by summarizing the questions which can be answered by each of the six alternative methods of data collection.

a. Longitudinal method. Historical description of age changes in a specific cohort under the assumption of stability (information about GLR cycling) or irreversible decrement; detection of species-specific age changes under the assumption of decrement with compensation.

b. Cross-sectional method. Detection of species-specific generation differences under the stability assumption; recommendations regarding treatment of individuals at different ages for immediate policy decisions under all assumptions; recommendations for long term policy decisions under the stability assumption.

c. Time-lag method. Detection of cultural change under the stability assumption.

d. Cohort-sequential method. Prediction of specific cohort performance detection of species-specific age changes, culture-specific changes and recommendations for long term policy decisions under the irreversible decrement assumption.

e. Time-sequential method. Historical account of age changes in a specific cohort, prediction of specific cohort performance and detection of culture-specific change under the decrement with compensation assumption. This method is also appropriate for information relevant to long term policy recommendation involving both cultural and age changes.

f. Cross-sequential method. Prediction of cohort-specific change under the stability assumption. Also appropriate for the separation of culture specific change into cohort and time of measurement components.

IV. Within Cohort (Longitudinal) Studies of Adult Development

We will now consider a number of problems in the implementation of descriptive research on adult development which are particularly relevant to studies concerned with describing or predicting age changes within generations, to which the longitudinal approach or the more generational sequential methods are applicable. Most of these problems have conceptual as well as practical implications for research design.

A. Development and Socio-cultural change

Except for those variables whose development is programmed by the biological characteristics of the organism, it is expected that environmental input is required to result in what we call developmental change. It is necessary here to distinguish different types of such input. First of all there is the input which occurs through the simple act of measurement itself, which may modify the response of the organism whether the experimenter so intends or not. Secondly, there is the experimental modification or treatment intervention programmed or selected by the investigator. Such treatment effects do appear frequently also in purely descriptive research, since we may have selected a population sample which is known to receive a treatment between measurement intervals which differs from that accorded to the parent population. Finally, there is the environmental impact occurring due to unspecified modifications in the environmental impact which the researcher has no control or of which he is not cognizant. In all three instances the treatment involved may equally effect organisms of any age, or we may have an age by treatment interaction; i.e. the treatment may effect organisms at one age but not at another.

The first two sources of treatment variance can be controlled by the traditional method of random assignment of subjects to different treatment or practice conditions (cf. Baltes and Goulet, 1971). No such assignment is possible, however, with respect to the third source, which involves socio-cultural change over which the experimenter has no control, illustrating the necessarily quasi-experimental nature (Campbell and Stanley, 1963) of any developmental research design. The dimension of socio-cultural change is of paramoung importance nn descriptive research on aging. It will be particularly bothersome in long-term studies since there is ample evidence that the attitudes of subjects about themselves and the experimental procedures will change, as will the attitudes of the experimenter which may differentially effect the outcome of his procedures.

If the researchers ingenuity and resources permit, he should then consider a mixed model which will permit the study of interactions between age level, experimental treatment effects and naturalistic treatment effects (the latter being attributed to socio-cultural change. Such model, however, requires that all age levels be examined at more than one time and under each of the experimental treatment conditions as well as a control. An empirical example of a design for the differentiation of age, time of measurement and treatment effects is reported elsewhere for the dimensions of behavioral rigidity (Schaie, 1971b).

The investigator conducting descriptive research may be willing to assume that there has been no systematic intervention for which he must control as in the case of experimental manipulation. He cannot safely ignore the confounding of the effects of measuring his variables and the general environmental impact in studies of aging. For those variables where environmental impact is expected (and this would include most variables of interest to the psychologist) the sequential designs must therefore be used to distinguish between age changes (qua maturational events and/or age specific though environmentally programmed behavior acquisitins) and the temporally unique generalized input from the environment.

If the investigator is interested primarily in population parameters, he may be able to beg the issue of measurement or practice effects by considering the independent random sampling design as the most general form of longitudinal inquiry. Here the investigator would take successive random samples from the same cohort at successive times of measurement, testing each subject only once. Specification of either a sampling with or without replacement should be made, although it is unlikely that there will be practical consequences of assuming one or the other sampling model. Further complications arise, of course, if individual age changes are to be observed which will require a repeated measurement design.

B. Effects of Repeated Measurement

In spite of the many sampling and sample maintenance problems discussed earlier, many investigators of within generation age changes have preferred which occurs through the simple act of measurement itself, which may modify the response of the organism whether the experimenter so intends or not. Secondly, there is the experimental modification or treatment intervention programmed or selected by the investigator. Such treatment effects do appear frequently also in purely descriptive research, since we may have selected a population sample which is known to receive a treatment between measurement intervals which differs from that accorded to the parent population. Finally, there is the environmental impact occurring due to unspecified modifications in the environmental impact which the researcher has no control or of which he is not cognizant. In all three instances the treatment involved may equally effect organisms of any age, or we may have an age by treatment interaction; i.e. the treatment may effect organisms at one age but not at another.

The first two sources of treatment variance can be controlled by the traditional method of random assignment of subjects to different treatment or practice conditions (cf. Baltes and Goulet, 1971). No such assignment is possible, however, with respect to the third source, which involves socio-cultural change over which the experimenter has no control, illustrating the necessarily quasi-experimental nature (Campbell and Stanley, 1963) of any developmental research design. The dimension of socio-cultural change is of paramoung importance nn descriptive research on aging. It will be particularly bothersome in long-term studies since there is ample evidence that the attitudes of subjects about themselves and the experimental procedures will change, as will the attitudes of the experimenter which may differentially effect the outcome of his procedures.

If the researchers ingenuity and resources permit, he should then consider a mixed model which will permit the study of interactions between age level, experimental treatment effects and naturalistic treatment effects (the latter being attributed to socio-cultural change. Such model, however, requires that all age levels be examined at more than one time and under each of the experimental treatment conditions as well as a control. An empirical example of a design for the differentiation of age, time of measurement and treatment effects is reported elsewhere for the dimensions of behavioral rigidity (Schaie, 1971b).

The investigator conducting descriptive research may be willing to assume that there has been no systematic intervention for which he must control as in the case of experimental manipulation. He cannot safely ignore the confounding of the effects of measuring his variables and the general environmental impact in studies of aging. For those variables where environmental impact is expected (and this would include most variables of interest to the psychologist) the sequential designs must therefore be used to distinguish between age changes (qua maturational events and/or age specific though environmentally programmed behavior acquisitins) and the temporally unique generalized input from the environment.

If the investigator is interested primarily in population parameters, he may be able to beg the issue of measurement or practice effects by considering the independent random sampling design as the most general form of longitudinal inquiry. Here the investigator would take successive random samples from the same cohort at successive times of measurement, testing each subject only once. Specification of either a sampling with or without replacement should be made, although it is unlikely that there will be practical consequences of assuming one or the other sampling model. Further complications arise, of course, if individual age changes are to be observed which will require a repeated measurement design.

B. Effects of Repeated Measurement

In spite of the many sampling and sample maintenance problems discussed earlier, many investigators of within generation age changes have preferred to conduct studies which involve the repeated measurement of the same subjects. Whether or not the intent has been to follow individual subjects over time, it has frequently been argued that in circumstances where experimental error is large and changes may be small, variability can be significantly reduced by use of the matched group design of which the longitudinal repeated measurement study is an application.

As indicated above, unless we can assume that there are no socio-cultural change effects upon our measurement variables and unless we know that the effect of one measurement upon subsequent measurements is zero, the results of a single sample repeated measurement study cannot be interpreted and such a research design, therefore, does not represent a permissible developmental research strategy. This prohibition does not apply to the use of repeated measurements in the sequential methods. However, it should be pointed out that the repeated measurement design may be less desirable than the independent random sampling approach for other reasons also. These are (1) lower efficiency, (2) misleading regression effects, (3) design complications in handling effects of practice, and (4) experimental mortality.

1. Comparative efficiency of repeated measurement designs.

Repeated measurement designs become efficient only when measures have high reliability or when it is possible to take measures either at many ages or many times of measurement. This is so because of the fact that for an equal number of observations the degrees available for the error term used to test mean differences will always be less for replicated than for independently derived measurements. The gain in sensitivity must, therefore, exceed the loss caused by testing a smaller number of subjects before the repeated measurement design will begin to pay off.

The above statement is not just a bias in favor of independent random sampling. It can be shown formally that the error term used for the test of significance of cohort differences will be based on the residual variation between subjects which in the cohort-sequential method will have N - (N/A) - C (A - 1) fewer degrees of freedom than would an independent random sampling design based on the same number of observations, where N is the number of observations, A the number of age levels and C is the number of cohort levels. Similarly the error term which must be used for testing the significance of cohort differences in the cross-sequential method will have N - (N/T) - C (T - 1 fewer degrees of freedom in the repeated measurement than in the comparable independent random sampling design. (T stands for time of measurement levels). The error term used to test age and age-by-cohort interaction in the cohort-sequential design will have N - (N/A) fewer degrees of freedom and the error term used to test time and time-by-cohort interaction in the cross-sequential design will be reduced by N - (N/T) degrees of freedom. It should also be noted that the time-sequential method with repeated measurement faces the problem that repeated measurements are not available for the first or last age level, and tha between subject sums of squares would, therefore, have to be computed upon the means for each subject's scores rather than upon the sums as would customarily be the case. In the latter case there would be N - (A + 2) (N/AT) fewer degre of freedom than in the independent random sampling design.

2. Misleading regression effects.

Orthodox test theory would suggest strong regression effects in any study where the same subjects are measured repeatedly there has been considerable controversy as to how change should be measured under such circumstances or whether it should be measured at all (Campbell and Stanley,1963; Cronbach and Furby,1970

The principal effect of regression consists in the assumed positive correlation between error or measurement and observed score, while the error scores of subjects on different occasions are uncorrelated. The effect of regression may not be too serious for the estimation of population parameters but it may lead to an artifactual shrinkage of differences between age levels, cohorts or times of measurement over long series of repeated observations. The effects of regression seem to be fatal, however, in studies contrasting ability levels. There seems strong evidence, for example, that evidence on differential development of intellectual function by ability level frequently reported in the literature (e.g. Miles and Miles, 1932; Owens, 1966; Riegel, Riegel and Meyer, 1967). By using appropriate time reversal procedures Baltes, Nesselroade, Schaie and Labouvie (1971) were able to show that ability-level related differences in ontogenetic patterns of adult intelligence found in a cross-sequential repeated measurement study (Schaie and Strother, 1968a) occurred due to statistical regression effects. MMreover, in agreement with Lord (1963) it was found that the regression effects were greatest for the least reliable measures. Time-reversal or similar techniques (Campbell and Stanley, 1963) are therefore strongly recommended whenever differences in developmental change as a function of initial performance level are to be investigated.

3. Design complications in handling effects of practice.

It has already been stated that the conventional longitudinal repeated measurement designs confounds age, environmental change and the effects of measurement or, practice. The sequential strategies separate the components of developmental change but in their most straightforward form also confound the effects of practice. They do permit a direct estimate of the effect of practice from the cohort-sequential repeated measurement design if age changes can be assumed to be zero. Similarly practice effects can be estimated in the cross-sequential repeated measurement design if time of measurement differences are assumed to be zero. If these assumptions do not hold, then practice will be confounded with either age or time of measurement effects.

It is possible to obtain independent estimates of practice effects under all assumptions, if the investigator is prepared to employ conjointly repeated measurement and independent random sampling designs. These estimates are obtained by examining all levels of the developmental components at two or more levels of practice. Minimal data to assess the effects of measurement or practice involve at least three test occasions for the cross-sequential and time sequential method and at least four test occasions for the cohort-sequential method. Figures 1 to 3 provide the necessary sampling plans and generalyzed analysis of variance models. Note that the estimates of practice effects are based on populations receiving different treatment and not upon within subject variance. The latter analysis is possible for data collections via the cross-sequential and cohort-sequential methods but will require even more complex sampling plans. Such designs may be needed when there are specific methodological reasons to estimate within subject practice effects. In general, however, the design here presented should provide appropriate strategies for accounting for the variance due to measurement effects.

4. Experimental mortality.

Perhaps the most serious limitation of the repeated measurement design is the inevitable non-random attrition or what Campbell and Stanley (1963) term experimental mortality. It may be profitable to distinguish here between two types of attrition. The first which may or may not be a function of the experimenter's skill in sample maintenance involves age differences in experimental mortality due to psychological and/or sociological reasons. For example, attrition caused by lack of interest, active refusal, change of residence or disappearance The second type of experimental mortality over which the experimenter has no influence involves age differences in sample attrition due to biological causes. Here we would include age-related subject loss due to physical illness and individual differences in longevity. Not all studies need be affected by this problem. However, in all those instances where there is a correlation between attrition and the variable of interest, findings on ontogenetic changes can be markedly affected. The problem seems particularly serious in the area of cognitive function since brighter subjects seem to have greater longevity (e.g., Blum, Jarvik and Clark, 1970; Jarvik and Falek, 1963; Riegel, Riegel and Meyer, 1968).

Studies of attrition caused by biological factors typically show that survivors have more positive attributes with respect to interests and attitudes (Riegel, Riegel and Meyer, 1968; Streib, 1966) but also higher social and educational levels (Rose, 1965; Streig, 1966). Lest it be thought that favorable selection of residual samples in longitudinal study be restricted to biological variables, there is also strong evidence that such favorable selection, at least with respect to intellectual level, occurs due to dropout from all causes (Baltes, Schaie and Nardi, 1971). Implications for the internal and external validity of longitudinal studies are discussed in further detail in the report of the latter study.

C. Changes in Factor Structure over Time

In addition to the problem that the reliable variance in developmental observations may be obscured by regression effects, we must further be concerned with the change in meaning also of the true component of any observed score. The most cautious approach would, of course, consist of relying solely on measures whose factorial structure and stability over time were well known. But measures of this nature are few and even fewer investigators in the past have paid much attention to the advantage of multivariate approaches to developmental problems (see also Schaie and Marquette, 1971). While the question of equivalence of test instruments is particularly pertinent when species-specific studies are examined which compare generation differences, it is the equivalence of factor structure for the same instrument which is at issue when we consider within cohort age-related variations of performance. The issue of equivalence of instruments should be important to the investigator of adult development only in those instances where he wishes to link the behavior of his subjects from childhood to adulthood. In the latter case equivalence of factor structure within samples over different instruments will also be at issue (e.g. Schaie and Cattell, 1971).

For within cohort studies there are three different problems related to the invariance of factor structure which should be considered. These concern the distinction between structural change in the relationship among dimensions and quantitative change in performance level on conceptionally identical dimensions (Baltes and Nesselroade, 1970; Cattell, 1969); the question of describing age functions for distinct and ideally orthogonal dimensions (Schaie, Rosenthal and Perlman, 1953; Botwinick, 1967; Reinert, 1970); and thirdly the disentanglement of ontogenetic and generation variance on factorially stable dimensions (Nesselroade, Schaie, and Baltes, 1971).

1. Structural change in factor structure

The first problem relates to the possibility that one could have systematic age changes on observables (due to observational artifacts) with concurrent stability of factor scores. That is we need to know not only whether the factorial structure of a given measurement set remains invariant over different measurement occasions, but also whether there are shifts in factor scores resulting from changes in differential weighting. The question of the invariance of factor structure normally involves the comparison of the correlation patterns among first occasion measures with the correlation patterns for subsequent measurement occasions. Evidence for structural change could take the form of different numbers of factors for different occasions and/or lack of invariance of the factor loading patterns. To protect against the possibility of reporting structural changes as a function of invariance of observational error it is recommended strongly that the zero order correlation matrices also be compared directly to test the hypothesis that their deviation from one another is no greater than chance (Schaie, 1958b). If such test fails to reject the null hypothesis, it would then be fair to conclude that changes in number of factors or factor loading patterns are chance variations and can be safely ignored.

Rejection of the null hypothesis with respect to equivalence of the zero order correlation matrices does not necessarily imply the acceptance of the alternate hypothesis of structural change. Before this outcome can be accepted (with the consegment conclusion that results of the study were uninterpretable with respect to quantitative change in performance level) it would be wise to rotate the factor matrices for the different measurement occasions to positions of maximum agreement by means of Meredith's procedure (1964) with subsequent transformation of the factor pattern to simple structure while maintaining invariance attained by the Meredith rotation (Nesselroade and Cattell, 1971).

2. Quantitative change in performance level

Once the factors which are invariant over time have been identified it then becomes possible and desirable to estimate factor scores by averaging corresponding.

first and subsequent occasion weights obtained from the factor pattern loadings (Harman, 1967). However, for a developmental study it is essential that observed score distributions be standardized jointly over all measurement occasions prior to estimating factor scores. The latter procedure is required in order to permit the appearance of between occasion mean differences.

The advantage of age functions plotted on the basis of factor scores over those using the observed raw measures should be obvious. We are comparing individuals on idealized variables whose conceptual meaning has been established to be invariant over time. Changes in mean performance level on factor scores can therefore validly be interpreted as change in behavior rather than as possible transformations in the meaning of the measurement variable as is the case when the observed variables are compared over time.

Of interest also will be the assessment of the stability of factor scores. Since the factor scores represent estimates of the reliable components of variance it may be argued that their magnitude should give an indication of the stability of the behavior rather than of the reliability of the measurement instrument.

3. <u>Ontogenetic and generation variance in factorially stable dimensions.</u>

The observation of factorial invariance in an adult population can, of course, not be taken as direct evidence of stability over age. Nevertheless it is most likely that both the irreversible decrement and the decrement with compensation models apply primarily to observable variables which are factorially complex, and whose complexity (i.e. factorial structure) changes over time. The stability model may therefore be most likely to apply to factorially pure and consequently structurally invariant variables. If this latter speculation should be correct, then it would follow that the cross-sequential would be most appropriate for the analysis of factor scores to distinguish between the proportions of variance accounting for mean differences in generational levels from that due to temporal variation in the GLR represented by the ontogenetic component of variance.

Nesselroade, Schaie and Baltes (1971) have recently reported an empirical example of this approach for various measures of cognitive behavior. This study clearly demonstrates that significant differences in mean factor scores can be found between generations as well as within generations. But the study also shows in a design which permits the appropriate comparison that the proportion of variance within generations (of major interest in longitudinal studies) is quite trivial and for no factor accounts for more than 5% of the total variance. Components of variance due to generational differences on the other hand account for at least one third of the total variance and may therefore be assumed to have practical as well as theoretical implications.

V. Between Cohort (Cross-sectional) Studies of Adult Development

The reader may by now have concluded that in the realm of adult behavior development little is to be gained from a description of age functions as such and that our interest might better be applied to the investigation of cultural change and its impact upon behavior found to occur in adults belonging to different generations at different points in historical time. It is hoped that he will have concluded further that while cross-sectional studies cannot yield age functions they are admirably useful for investigating inter-generational differences. Both the traditional cross-sectional method and its more general form, the time sequential method, can be used for this purpose. But studies of inter-generational differences have their problems also. Three different issues will here be considered in this context. The first addresses itself to the question as to what data on generational differences can be obtained from a single cross-sectional study as compared with an analysis of differences between successive cross-sectional studies, the time-sequential method. Secondly, I would like to point to possible designs involving both prospective and retrospective data, and lastly it will be necessary to raise the question whether age or cohort-appropriate measurement devices should be used when we are interested in assessing generation differences.

A. Comparisons Within and Between Cross-sectional Studies

If the student of adult development is willing to accept the stability model he can then utilize the single time of measurement cross-sectional study to describe differences in behavior between successive generations. This means that much of the psychological literature on age differences can once again become useful if we engage in the massive effort of relabling tables and graphs by substituting the subjects' year of birth (identifying their cohort membership) for the age designation. That is, whenever we are willing to assume that a variable does not show any significant age changes over the adult period we can interpret the results of cross-sectional studies as direct means of describing inter-generational differences. Taken from this point of view it now becomes necessary to conduct a systematic re-review of the literature, since failure to replicate findings of performance level for given ages is no longer critical. What we are looking for now would be the replicability of findings for given cohorts.

But what happens in the event that the assumption of stability of adult behavior were false? Surprisingly enough, no serious consequences would result in those instances where cross-sectional data are appropriate in the first place Since cross-sectional data would normally be applied only to problems requiring judgment about population differences at the present point in time, it would not really be crucial if one of the components of the observed difference between successive cohort were due to change in maturational level. On the other hand the failure of the stability assumption would be quite critical where the cross-sectional model is used to detect generation differences for the purpose of long-range forward planning.

Studies directed towards generating data for forward planning assume, of course, that inter-generational differences detected at the present time are likely to shift. If the reader will recall our analysis of the principal models accounting for adult development, he will note that this assumption is most compatible with the decrement with compensation model. That is, we want to know to what extent does shift in environmental input effect level of performance at different ages. The comparisons provided by the time-sequential method, do indeed provide the necessary information. That is, we can here isolate the components of variance due to change in maturational level, from those due to change in environmental input, as well as from the shift in inter-generational differences which would be represented by the interaction term. That is, if we replicate a cross-sectional study over several measurement occasions, we will find that age differences observed at one measurement point which are actually generation differences will tend to disappear when averaged over many measurement points. But if they are indeed differences in maturational level, they will remain, no matter how many replications are taken. Likewise, environmental input will be assessed by comparing the difference between measurement occasions over many age-generation levels, and it becomes possible after a few replications to plot projected trends which may be useful for long range predictions.

It should be noted, though, that comparison of cross-sectional studies based on independent samples can only yield information on the stability of performance levels within generations under the assumption of the stability model. Other models would require replication of the same cross-sectional study by drawing successive independent samples covering the same age range from the same parent population to draw meaningful conclusions.

B. Comparison by Means of Prospective and Retrospective Estimates

Although the one-shot cross-sectional approach has been faulted ad nauseam, it remains obviously attratice to the researcher (such as the graduate student) who has only limited time to complete a given project. Serious attention has therefore been given to the possibility of generating data covering the adult life span by obtaining retrospective or prospective information over many points of the respondents experience but collected at one point in time (e.g. Bell, 1960; Lehr, 1967; Yarrow, Campbell and Burton, 1970). Leaving aside the serious questions with respect to the reliability and validity of information thus gathered, it should be noted that such studies suffer from the general problems of the longitudinal repeated measurement model, although in an asymmetric fashion. That is, in the retrospective report of behavior, each earlier episode affected made by the report of more recent episodes as well as the present condition of the respondent. On the other hand in prospective reports, each later projected episode will be affected by reports about earlier episodes. Whether or not the sequential methods are applicable to the study of retrospective and prospective data remains to be investigated. One major prob here, would be the necessity of comparing cohorts with unequal saturation of relative past and future life intervals at different age levels. An attempt at utilizing such models has been made by Back and Bourgue (1970) in a study of life satisfaction.

While there are difficulties in using prospective and retrospective methods for the study of actual behavior, these methods seem to be quite suitable for the consideration of perceived age and cultural changes. This area, which has received considerable recent interest, is bound to expand further if Neugarten (1969) were to be correct in her assertion that role perception associated with age level is indeed one of the most important determinant of ontogenetic change. Since role perception of any age level must be viewed from a constant point in historical time, the cross-sectional approach is quite appropriate here, and the time-sequential method would be used to test the stability of such role perceptions. (See also Ahammer and Baltes, 1970; Bekker & Taylor, 1966; Britton and Britton, 1969).

C. Age-appropriate vs. cohort appropriate test instruments

Whenever comparisons are to be made between cohorts we must concern ourselves with the question whether or not test instruments are equally appropriate for members of different cohorts. This issue has been studied for some time with respect to tests thought appropriate for different age levels. For example, forms suitable at different ages have been developed for the Wechsler intelligence tests or for the factored series of personality tests developed by Cattell and his associates. The fact that a given test is or is not appropriate for a given age level does not, however, assure us that it will remain appropriate for that age level. That is, because of changes in the content of the culture, at least verbal tests, may become inappropriate for a given age level. In this sense it is conceivable that the original Wechsler-Bellevue may be a more valid test for older adults than is its more recent revision. And that revision may now be discriminating against twenty year olds because the cohort appropriateness of its content has shifted.

Perhaps the most serious work bearing on the issue of cohort appropriateness of test material, crucial for adequate comparison of the membership of different generations, is being conducted by Monge (1971) and his associates. That group of investigators classified words in terms of their entry into the accepted vocabulary as demonstrated by appropriate dictionary search. They were able to show that the cohort word familiarity gradient could be readily manipulated for slang items and other words depending upon the entry of the term into active vocabulary. Thus, while twenty and thirty year olds performed significantly better on items entering the language after 1960, it was found that 40 to 60 year olds performed significantly better on items entering the language in the late nineteen twenties. It appears then that differences in verbal performance between cohorts typically reported from cross-sectional studies could well be artifacts related to the cohort appropriateness of the test instruments used.

VI. Some Comments on Research Priorities in the Study of Adult Development

I prefaced my discussion of methodological issues in descriptive research on adult development, by saying that the time was not yet ripe for laboratory manipulative investigation because we need to take another look at what is to be manipulated. This does not mean, however, that we should ignore the possibilities offered by experiments of nature nor that we should revert to an aimless search of evidence which differentiates the behavior of the old from that of the young. I would like to suggest, however, that high priority be given to a hypothetico-deductive approach which specifies the kind of models according to which adult development is expected to occur and then to assemble data which could support or discredit such models. This means that we should not reject the use of cross-sectional approaches out of hand, but rather, by means of suitable replication, test the generalizations possible for such data. I would further suggest that studies of extreme groups should be encouraged because they are likely to yield quicker payoff in terms of developmental laws regarding adult development then the search for representative data which as I have demonstrated may often be irrelevant or impossible to attain. Where population estimates are sought, again it would seem appropriate to give much thought to the purpose for which data are to be collected. Finally, I would argue that the emphasis for students of adult development, at least for psychological variables, must be directed towards intergenerational differences and cultural change as more meaningful parameters than that of chronological age. Only thus can we escape the constraints that undue preoccupation with age as an independent variable has imposed upon our endeavors to understand and predict the course of adult development.

References

Ahammer, I.M., & Baltes, P.B. Desirability judgments of adolescents, adults, and older people: Self vs. hetero-perceptions. Paper presented at meeting of the American Psychological Association. Miami, 1970.

Back, K. W., & Bourgue, L.B. Life graphs: Aging and cohort effects. Journal of Gerontology, 1970, 25, 249-255.

Baltes, P. B. & Goulet, L.R. Status and issues of a life-span developmental psychology. In L.R. Goulet & P.B. Baltes (Eds.), Life-span developmental psychology: research and theory. New York: Academic Press, 1970, Pp. 149 191.

Baltes, P.B., & Goulet, L.R. Exploration of developmental parameters by manipulation and simulation of age differences in behavior. Human Development, 1971, in press.

Baltes, P.B., & Nesselroade, J.R. Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. Developmental Psychology, 1971, in press.

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 adult intelligence. Developmental Psychology, 1971 in press.

Baltes, P.B., Schaie, K.W., & Nardi, A.H. Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. Developmental Psychology, 1971, in press.

Bekker, L.D., & Taylor, C. Attitudes toward the aged in a multi-generational sample. Journal of Gerontology, 1966, 21, 115-118.

Bell, R.Q. Retrospective and prospective views of early personality development Merrill-Palmer Quarterly, 1960, 6, 131-144.

Birren, J.E. (Ed.) Handbook of aging and the individual: psychological and biological aspects. Chicago: University of Chicago Press, 1959.

Blum, J.E., Jarvik, L.F., & Clark, E.T. Rate of change on selective tests of intelligence: A twenty-year longitudinal study of aging. Journal of Gerontology, 1970, 25, 171-176.

Botwinick, J. Cognitive processes in maturity and old age. New York: Springer, 1967.

Britton, J.O., & Britton, J.H. Discrimination of age by preschool children. Journal of Gerontology, 1969, 24, 457-460.

Buech, B.U., & Schaie, K.W. Generational vs. ontogenetic components of change: A second follow-up. Paper to be presented as part of a symposium on "Intellectual Development in Adulthood and Old Age", American Psychological Association, Washington, 1971.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 171-246.

Cattell, R.B. Theory of fluid and chrystallized intelligence: a critical exper ment. Journal of Educational Psychology, 1963, 54, 1-22.

Cattell, R.B. Comparing factor trait and state scores across ages and cultures. Journal of Gerentology, 1969, 24, 348-360.

Cronbach, L.J., & Furby, L. How we should measure "change" - or should we? Psychological Bulletin, 1970, 74, 68-80.

Davies, D.F. Mortality and morbidity statistics. I. Limitations of approaches to rates of aging. Journal of Gerentology, 1954, 9, 186-195.

Harman, H.H. Modern factor analysis. Chicago University of Chicago Press, 19[illegible]

Jarvik, L.F., & Falek, A. Intellectual stability and survival in the ages. Journal of Gerentology, 1963, 18, 173-176.

Lehr, U. Attitudes towards the future in old age. Human Development, 1967, 10, 230-238.

Lord, F.M. Elementary models for measuring change. In C.W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963

Meredith, W. Rotation to achieve factorial invariance. Psychometrika, 1964, 29, 187-206.

Monge, R.F. Experimental tests from the Syracuse Adult Development Study. Personal Communication, 1971.

Nesselroade, J.R., & Cattell, R.B. Examination of short-term invariance and stability of personality factor structure. Unpublished manuscript, West Virginia University, 1970.

Nesselroade, J.R., Schaie, K.W., & Baltes, P.B. Ontogenetic and generational components of structural and quantitative change in adult cognitive behavior. Journal of Gerentology, 1971, in press.

Neugarten, B.L. Continuities and discontinuities of psychological issues into adult life. Human Development, 1969, 12, 121-130.

Owens, W.A. Age and mental ability: A second adult follow-up. Journal of Educational Psychology, 1966, 57, 311-325.

Platt, L. & Parkes, A.S. Social and genetic influences on life and death. New York: Plenum Press, 1967.

Reinert, G. Comparative factor analtic studies of intelligence throughout the human life span. In L.R. Goulet & P.B. Baltes (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970. Pp. 467-484.

Riegel, K. F., Riegel, R.M., & Meyer, G. A study of the drop-out rates in longitudinal research on aging and the prediction of death. Journal of Personality and Social Psychology, 1967, 4, 342-348.

Riegel, K.F., Riegel, R.M., & Meyer, G. The prediction of retest resisters in research on aging. Journal of Gerontology, 1968, 23, 370-374.

Rose, C.L. Representativeness of volunteer subjects in a longitudinal aging study. Human Development, 1965, 8, 152-156.

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

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. In L. R. Goulet & P. B. Baltes (Eds.), Life-span developmental psychology: New York: Academic Press, 1970. Pp. 485-507.

Schaie, K.W. Can the longitudinal method be applied to the psychological study of human development. Paper to be presented at the First Symposium of the the International Society for the Study of Behavioral Development. Nijmegen, 1971. (a)

Schaie, K.W. Cultural change and repeated assessment in the study of the adult personality. Paper to be presented as part of a symposium on "Personality Assessment of the Aged", American Psychological Associaton, Washington, 1971. (b)

Schaie, K.W., & Cattell, R.B. Upper and lower boundary alignment of personality factors measured by the Child Personality Questionnaire. Unpublished manuscript. West Virginia University, 1971.

Schaie, K.W., & Marquette, B.W. Personality in maturity and old age. In R.M. Dreger (Ed.), Multivariate personality research: Contributions to the understanding of personality in honor of R. B. Cattell. Baton Rouge, La.: Claitor, 1971.

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

Schaie, K.W., & Strother, C.R. A cross-sequential study of age changes in cognitive behavior. Psychological Bulletin, 1968, 70, 671-680. (a)

Schaie, K.W., & Strother, C.R. The effects of time and cohort differences on the interpretation of age changes in cognitive behavior. Multivariate Behavioral Research, 1968, 3, 259-294. (b)

Schaie, K.W., & Strother, C.R. Cognitive and personality variables in college graduates of advanced age. In G.A. Talland (Ed.), Human aging and behavior recent advances in research and theory. New York: Academic Press, 1968. Pp. 281-308. (c)

Streib, G.F. Participants and drop-outs in a longitudinal study. Journal of Gerontology, 1966, 21, 200-201.

Wohlwill, J.F. Methodology and research strategy in the study of developmental change. In L.R. Goulet & P.B. Baltes (Eds.), Life-span developmental psychology: research and theory. New York: Academic Press, 1970. Pp. 149-191.

Yarrow, M.R., Campbell, J.D., & Burton, R.V. Recollections of childhood: A study of the retrospective method. Monographs of the Society for Child Development, 1970, 35, 5 (Whole No. 138).

FIGURE 1

Sampling Plan and Generalized Analysis of Variance Model
for the Cohort-Sequential Method
Controlled for the Effects of Practice

a. Sampling Plan*

Time of Test	Samples Tested
T 0	Random half of group C_iA_m
T 1	All of C_iA_m; random halves of groups C_iA_n and C_jA_m
T 2	All of C_iA_n; all of C_jA_m; random half of group C_jA_n
T 3	All of C_jA_n

*Scores obtained on the first test for each sample are disregarded. Only the second set of scores for the random halves tested initially and the single set of scores available for the random halves without practice enter into the analysis.

b. Generalized Analysis of Variance MOdel

Source of Variation	Degrees of Freedom
Between Cohorts (C)	C - 1
Between Ages (A)	A - 1
Between practice levels (P)	P - 1
Cohort x Age interaction	(C - 1) (A - 1)
Cohort x Practice interaction	(C - 1) (P - 1)
Age x Practice interaction	(A - 1) (P - 1)
Cohort x Age x Practice interaction	(C - 1) (A - 1) (P - 1)
Error	N - (C) (A) (P)
Total Variance	N - 1

FIGURE 2

Sampling Plan and Generalized Analysis of Variance Model
for the Time-Sequential Method
Controlled for the Effects of Practice

a. Sampling Plan*

Time of Test	Samples Tested
T 0	Random halves of samples A_mT_k and A_nT_k
T 1	All of A_mT_k and A_nT_k; random halves of A_mT_1 and A_nT_1
T 2	All of A_mT_1 and A_nT_1

*Scores obtained on the first test for each sample are disregarded in this analysis. Thus the second set of scores for the random half tested initially and the only set of scores for the random half without practice enter the analysis.

b. Generalized Analysis of Variance Model

Source of Variation	Degrees of Freedom
Between Times (T)	T - 1
Between Ages (A)	A - 1
Between Practice Levels (P)	P - 1
Time x Age interaction	(T - 1) (A - 1)
Time x Practice interaction	(T - 1) (P - 1)
Age x Practice interaction	(A - 1) (P - 1)
Time x Age x Practice interaction	(T - 1) (A - 1) (P - 1)
Error	N - (T) (A) (P)
Total variation	N - 1

FIGURE 3

Sampling Plan and Generalized Analysis of Variance Model for the Cross-Sequential Model Controlled for Effects of Practice

a. Sampling Plan*

Time of Test	Samples Tested
T 0	Random halves of samples C_iT_k and C_jT_k
T 1	All of C_iT_k and C_jT_k; random halves of C_iT_l and C_jT_l
T 2	All of C_iT_l and C_jT_l

*Scores obtained on the first test for each sample are disregarded. The second set of scores for the random halves tested initially and the only set of scores for the random halves without practice enter the analysis.

b. Generalized Analysis of Variance Model

Source of Variation	Degrees of Freedom
Between Cohorts (C)	C - 1
Between Times (T)	T - 1
Between Practice Levels (P)	P - 1
Cohort x Time interaction	(C - 1) (T - 1)
Cohort x Practice interaction	(C - 1) (P - 1)
Time x Practice interaction	(T - 1) (P - 1)
Cohort x Time x Practice interaction	(C - 1) (T - 1) (P -1)
Error	N - (C) (T) (P)
Total variation	N - 1