Revised Manuscript of

Workshop Presentation on

R E S E A R C H   M E T H O D O L O G Y
I N   T H E   S T U D Y   O F
H U M A N   A G I N G

Presented at

ISSBD China Satellite Conference

July 1987

K. Warner Schaie

Evan Pugh Professor of

Human Development and Psychology

Pennsylvania State University

University Park, Pennsylvania

USA

RESEARCH METHODOLOGY IN THE STUDY OF HUMAN AGING

INTRODUCTION

The purpose of the workshop presentation reported in this chapter was to assist researchers in developmental psychology to enhance their skills in the formulation of research questions, design of studies, and selection of appropriate methods of analysis for the examination of problems in the study of human aging. Although this workshop does not offer instruction in statistical methods, we will attempt to indicate how certain methods are specifically relevant for the design and analysis of data sets on psychological aging. As part of the workshop presentation, we will therefore briefly consider at the conceptual level the applicability of methods such as complex ANOVA designs (including sequential strategies), Cattell's data box and certain correlational methods. Substantive examples given in this chapter draw heavily upon my own work on cognitive development in adulthood.

This workshop presentation summarizes much of the material that I normally teach at the Pennsylvania State University during a semester-long graduate seminar. In order to compress approximately 20 hours of lectures within a 5-hour workshop, this material had to be presented in a much more sketchy and summary nature then I would have preferred if more time had been available. At the end of this chapter, the reader will find a bibliography that contains some general reference works as well as specific sources for the topics covered. In that bibliography I have also tried to provide examples of empirical studies in which the methods that I describe have been applied, and that might serve as models for those of you who have not had previous exposure to these techniques. All of this material should be reasonably comprehensible to persons who have had some coursework on the general linear model (more specifically Analysis of Variance and Multiple regression), an introduction to tests and measurements, and an introduction to developmental psychology.

This presentation is divided into four topical units. I begin by briefly discussing the special methodological requirements of <u>Adult Development as a Field of Inquiry</u>. This material immediately leads into the second topic, that of <u>Problems of measurement in the Developmental Sciences</u>. Here I deal with issues of scaling and of selecting units of measurement that may be quite familiar to you, possibly from a course in educational tests and measurements. But most of our attention here is given to a discussion of questions of reliability and validity as they affect studies involving the age variable. Our third unit will cover <u>Experimental Manipulation of the Age Variable</u>. Here I will first consider age-comparative studies that involve laboratory paradigms to manipulate both organismic and experiential variables. I will then consider an example of a within-age-group study that introduces experimental manipulations in an attempt to reverse age-related decline. In the fourth unit I turn to descriptive studies that use Quasi-experimental Designs and in particular exercise the reader in some detail through the circumstances under which Longitudinal or Cross-sectional methods may be most appropriate. As part of this topic methods are discussed that control for the threats to internal validity identified in Topic 2. A final comment deals with the interpretation of the significance and magnitude of group differences found in the aging literature.

TOPIC ONE: ADULT DEVELOPMENT AS A FIELD OF INQUIRY

Some Definitions and Historical Perspectives

In order to place the materials of this chapter in their proper context I want to begin by providing some definitions and historical perspectives. I shall comment briefly on the history of gerontology and geropsychology, provide some definitions on aging and development, and introduce several models accounting for differential patterns of aging.

The Development of Gerontology and Geropsychology

Why should psychologists be interested in studying human aging? There have always been a few people who have gotten to be very old, but is a phenomenon of the 20th century, that the number of people who have lived to be old is getting to be a large number. There are two imnportant reasons for this. The first one is that we have virtually in every country succeeded in reducing infant mortality. The second reason is that we have conquered infectious diseases, so that more people live until they reach the time in life when they die from the diseasse of old age rather than diseases that can occur by chance at any time during life. As a consequence there have been dramatic changes in the world's population with respect to age, changes which can be expected to accelerate over the next 50 years or so. The data I will refer to come from the United States census

Demographic changes

Increases in the proportion of elderly in the population. Figure 1 shows that in the United States in 1985 half of the population was under age 31 and half over that age. This figure also shows what we can expect by the year 2050. In that year, half of all Americans will be over age 41, and the age distribution will have become a lot more rectangular. Note in particular that there will be many more older women than older men. In fact, many of the issues of the very old age turn out to be women's issues.

Insert Figure 1 about here

Increases in longevity. Figure 2 shows some data that indicate the number of years remaining at different ages. These data come from the 1980 United States census. In 1980, the average number of years a person could expect to live in America was about 75 years. But, notice that those people who lived to age 75, could expect on average to live another 10-11 years; so that one-half of people at age 75 could expect to be 85 years old. These facts have enormous impact on the structure of our society, on all societies.

Insert Figure 2 about here

Development of scientific interest

Explosion of amount of scientific information. Because of the above-mentioned demographic changes, many scientists have become increasingly interested in the facts of aging, in both physiological and psychological realms. This development can be demonstrated by charting the number of publications appearing in journals that have content in the psychology of aging from 1880 to 1979 (vsee Figure 3). Until 1920, there were not enough publications to show up on a scale with units of 100. There were perhaps one or two publications a year. But from then on you can see the steep rise. By 1950 behavioral scientists had begun to realize, and by 1980 certainly had discovered, that aging was an important topic in psychology.

Insert Figure 3 about here

Emergence from studies of child development. It is interesting to note that the psychology of aging, in part, developed out of the study of children. In some of the longitudinal studies of children, the scientists began to discover that children do not stay children, but they become adults. And eventually even become old. If a great deal of information is already avalaiblable on some individuals, it is very profitable to collect even more, because you can then really understand what happens to individual lives. In the United States, a number of scientists who started out as child psychologists became interested in aging as they themselves became older. For example, one of the famous pioneers in developmental psychology G. Stanley Hall, studied children most of his life, but his last important scientific contribution is a book called "Senescence."

Aging and Development

Adults as agents of their own development

For those readers who may have been primarily concerned with the study of children it may be important to draw some distinctions between aging and development. Children follow both biologically and societally, a fairly circumscribed path of development. For example, among newborn children, it is impossible to find anyone who is capable of talking. By about two years later, all but a few very retarded children are able to engage in some form of verbal communication. We know that in childhood, individual differences are relatively narrow, but once you adulthood is reached, individual paths of development can become very different. That is so because adults have much more influence over their own development than do children. For example, adults can influence their health. They can decide, for example, whether they want to smoke or not. They can decide whether they want to spend their leisure time sitting in front of a television set, or engage in educational and stimulating activities. They can decide when they finish school to cease all further learning, or they can spend the remainder of their lives gaining further knowledge. That is why in adulthood, we expect to find much greater individual variability than in childhood. Some adults will decline very early, while others will go on growing until they die.

Factors effecting adult development

Adult development is mediated by several distinct types of factors some which are constraints over which individuals do not have any control, while others can be shaped by the individual's own decisions. For example, we are limited by the fact that as living organisms we are subject to certain species-specific biological constraints imposed upon us. There are environmental factors which are a function of the nature of the society that we live in, and of course there is always the interaction between the two. We can usefully distinguish three major influences on development: normatively age graded, history graded, and non-normative (see Figure 4).

Insert Figure 4 about here

Age graded influences. Some examples of "normatively age graded" influences follow: We attain puberty at a certain age, we go to school at a certain age, we generally expect to get married over a very narrow age range. And there comes a time when our biological equipment ceases to function properly. Past mid-life our eyes and ears start failing and our energy level declines. There is very little we can do about these influences.

History graded influences. An example of normative historical influences would be the fact that 50 years ago, a college education would have been available only to young people whose parents had a lot of money. Historical events have changed the availability of college education dramatically. There are other important historical influences, such as the consequences of wars. The development of children who grow up during a period of dislocation and scarcities caused by a war may be affected negatively, something which would not happen if these children had grwon up during a peaceful historical period. Historical influences also markedly effect the information available to individuals which may shape their behavior and development. For example, today we can know almost instantly about events occurring on the other side of the world. Many of us are able to use computers to access information in minutes which some 10 to 20 years ago would have taken years assemble. For developmental scientests, therefore it is important to specify not only at what age individuals are studied, but also at what time in history that information has been gathered.

Non-normative influences. Not all influences, however, are determined by biology and hisotry. There are indeed manys "non-normative influences" that have to do with individual life circumstances over which many people have a certain amount of control. For example, it is possible for an individual to decide not to get married until that person is 40 year old. It is possible for a young person to decide to choose one profession rather than another. For example, the personal decision of many readers of this chapter to become psychology students has the consquence of making their life quite different from what it would be if they had decided to become accountants. Throughout life, most persons encounter many individual experiences that have little to do with their age at that time or with the particular historical period they live in. Instead, these influences occur as a consequence of individual life events, and the manner in which individuals respond to such events.

Models of aging

Although there is no question that the period of childhood is associated with growth and increment in most behavioral functions, there is far less agreement upon the modal course of development during adulthood. At least three fundamentally different models must be considered as plausible (see Figure 5). One very common model of aging that is sometimes presented as prevailing for all human functions is characterized by the attainment of an asymptote in young adulthood with rapidly accelerating decline. We might call this the "irreverrible decrement model." This model may be appropriate for many physiological functions, as well as molecular behaviors such as simple reaction time. For other more complex phenomena, however, particularly for variables such as skill in verbal communication, a much more reasonable model is to assume that once an asymptote is reached, most adults will maintain that level as long as they remain in reasonably good health. This second scenario for adult development is what we would call an "adult stability model". A third model, which we have called "decrement with compensation" applies even more frequently. For example, most people's eyesight declines with age, there is indeed age-related decrement in visual accuity, but most individuals can compensate for this change in visual accuity by beingg fitted with appropriate eye-glasses. Similarly, energy level is known to declinw with age, but most older persons can adequatly compensate for their lower energy level by taking a little bit more time for everything they do. For some variables we might even expect a "linear increment model" to apply; hopefully, that would be true for something that we call "wisdom." In fact, we would argue that for the attainment of qualities like wisdom, advanced age is a necessary condition, even though it is not a sufficient condition.

Insert Figure 5 about here

The Formulation of Research Questions
in the Study of Adult Development

Temporal trajectory of the research question

Ayone proposing to do research on human development needs to be explicit in defining the temporal aspects of their research question. There are two fundamentally different questions that may be asked: The first one is the issue of concurrent status or age differences, the second is concerned with understanding age-related change within individuals.

Concurrent status (age difference research). Here we are interested in asking the question whether there is a difference between groups of people at different ages. For example, we want to know whether the behavior of children differs as a function of age. Likewise we may want to know whether there are differences between people at certain ages, for example, if we are interested in the issue of what would be a good average age for people to retire. In all of these instances we are not concerned with the previous or future status of individuals, but we simply want to note how people classifeid by chronological age differe on other variables of interest. These are very different questions, than those we would ask if our interest is focused upon establishing the facts of developmental progression.

Developmental progression (age change research). The distinction between age changes and age differences will receive much more attention later in this chapter. For the moment, it is important for the reader to understand that if the research question is one that involves change over time, then it would be irrelevant to compare people of different ages at one point in time. For example, I am going to be 60 next year. If you would like to predict what I will be like when I am 70, it would not help you to investigate people who are 70 years old today. Because when they were 60 years old, they were probably not like what I am today. This is a particularly critical issue in a society that experiences very rapid changes. Relying upon age difference research to answer questions that involve change over time therefor can lead to misleading conclusiongs which are sometimes called the "cross-sectional fallacy" very dramatically in America. I suspect that the "cross-sectional fgallacy" may pose even more serious problems for social and behavior scientists in China, because your society is changing so much more rapidly than has been the case in America.

Ideological goals of research questions

No research question is ever completely free of ideology. It is important, therefore, when you plan research to decide why you are doing the research and what it is that you want to know. We can broadly classify research questions of interest to developmental psychologists into four groups: Understanding of basic processes, gathering Descriptive data for purposes of policy formation, Prediction of the future status of individuals, and Evaluating othe effects of develelopmental interventions.

Understanding of basic processes. Many readers may perhaps be primarily interested in understanding very basic psychological processes. But the decision to study a particular process is usually not arbitrary, but rather is based on some informal theoretical considerations that link the process of interest to eventual practical consequences. The chain of evidence from the specific laboratory experiment to eventual applications of societal interest may be incomplete and extend over a long period of time. Nevertheless, thoughtful researchers will always be able to specify why the particular process they are studying is worthy of the attention given to it.

Gathering descriptive data for purposes of policy formation. A much more immediately recognizable goal for some research activities may be to gather descriptive data that can be used to lay a basis for policy formation. For example, we may wish to obtain data relevant to questions such as: At what age can people benefit from education? At what age should people retire? At what age can people largely be trusted to take care of their own affairs? Normative data may also be needed to determine cut-off points for access of different levels of education, or for the provision or denial of a variety of social services and opportunities that on individual differences, and in such instances as were optimal allocation of scarce resources is required.

Prediction of future status. Another reason for a particular research study may be because we wish to predict someone's future status. For example, will participation in a workshop such as the one that led to this volume help psychology students to become better researchers? A predictive study would involve collecting data, say some 10 years from now to see if participants of the workshop became more productive and better researchers than a group of their contempraries, who did not participate in the workshop.

Evaluation of interventions. The example just given above is also pertinent to the issue that many of our research studies involve the evaluation of intervention in development. For example, are some child rearing techniques better than others? Are some environments more intellectually stimulating than others? If we were to change the nature of some of the characteristics of a work setting, or change in environment of a factory, what would this do to the development of those affected by the change? Note that intervention studies may involve interventions that are specifically programmed for research purposes, or they may involve the monitoring of serendipitously occuring soicetal interventions, that can be studied to answer the researchers question.

What I am arguing then is that it is very important from the very beginning of any study to understand the specific purpose and goal of one's research. As we will see, it is the nature of the research question which will directly effect the choice of the most appropriate design, as well as indicate the nature of the the controls and analyses that will be required to obtain unambiguous andweres to the research question.

4Operationalization of Research Question

Identification of dependent and independent variables

When my students first come to me and want to do a research project, they generally want to do research on a very big or global topic, like the study of wisdom, or why people want to retire, etc. I immediately inform them that I do not know how to do research global topics; I tell them they have to think a little smaller. They need to take some dependent variable, say something like performance in associative memory, or perhaps some personality trait like flexibility, something smaller like that and then think about what the independent variables might be which may affect their dependent variable. Although in theory every variable can affect any other variable, the immediate decision that must be faced by every researcher is to decide as to what specific independent variables are to be studiedstudy. Whatever the choice of the research problem, one of the first necessary steps is to specifi what is the dependent variable, and what are the explanatory or independent variables. That is how any doable research project must begin.

<u>Selection of samples</u>

In aging research, at least in the United States, the naive approach well represented in the journal literature, is to compare two groups of people, one say age 20, and another group of people, say age 70. Typically, the young adults are college students, and the elders are retired people in some recreational group for old people. But if we compare groups such as these, we have problems, because the young people will be much better educated than the older people, the young people will be much healthier than the old people, and if we find a difference between the two groups, it may have nothing to do with age, but instead be due to effects of education or health. If a researcher is really interested in the age variable, then it is necessary to compare groups of people who are reasonably similar in other characteristics. As psychologists we must learn from our colleagues in sociology, who are much more careful in picking what might be called representative samples. Moreover, when the intent of a study is to determine how people differ across age, it is not sufficient simply to study young and old people. In studies of adulthood, it is consequently important to sample at least several points across the lifespan.

<u>Selection of age- and cohort- relevant research instruments</u>

Not all psychological tests and procedures or questionnaires are equally valid for people at different ages. For example, if you were to study the personality trait of social responsibility, and you are not careful in constructing your questionnaire, you might find that you have some items which are very appropriate for young adults, but totally inappropriate for old adults. This problem applies also to intelligence tests. An intelligence test that was constructed, say 20 years ago, would be much easier for people today than it was 20 years ago. Pilot studies are generally required when a test or questionnaire developed on a sample of one age is to be applied to study other age groups, or when instruments that were developed at much earlier points in time or in other societal settings are to be used.

Descriptive vs Hypothesis Testing Research

Distinction between experiments and quasi-experiments

An important distinction in the design of developmental research studies is the contrast between experiments and quasi-experiments. The major characteristic of a true experiment is the fact that it is possible to assign subjects randomly to different conditions, such that one can be sure that an experimental and a control group will be equivalent in all characteristics except for the treatment given to the experimental group. The problem with aging research, however, is that it is not possible to assign people randomly to a particular age. That is why we typically do quasi-experiments in aging studies. In quasi-experiments we do not randomly assign, but rather we try to match subjects across experimental conditions; that is, we try to composed equivalent control groups. Within the group of quasi-experimental designs available to us we still need to distinguish between studies that are primarily descriptive and those that are hypothessis-testing in nature.

Descriptive research

The purpose of such studies is to obtain information on differences in level of performance between groups of different ages (age-comparative research) or to track groups of individuals over time in order to determine developmental progressions or to describe differential patterns of development attributable to alternate antecedent conditions.

Importance of sampling plan. I have already mentioned that one of the most crucial issues in descriptive research is the formation of adequate sampling plans. In age-comparative (cross-sectional) studies the major concern is with the composition of samples that are equivalent with respect to all characteristics other than those measured as dependent variables. In studies across time (longitudinal) in which the same individuals are followed, one must also be concerned about what will happen to one's sample as time passes. Some of ther young adult participants may be lost because they are sent to some far away jobs. Some older participants may die, or may become too sick to be studied. If the researcher does not maintain good relations with subjects in a longitudinal study, some of them may decide that they do not want to remain in your study. Longitudinal studies consequently require a lot more effort than a single point study where data are collected from the subjects and there is no further need to ever see them again.

Statistical vs. experimental controls. We must also concern ourselves with the question to what extent can we use experimental controls and to what extent must we use statistical controls. In may instances there is no choice but to use statistical controls. For example, we it may not be possible to match different age groups in terms of their health, or in terms of their level of education in which case it may be desirable to make adjustments by means of statistical controls. Approaches available for this purpose include not only the Analysis of Covariance, but also other methods for creating base-free measures that adjust for initial group differences at entry into a study. It should be noted though, that equating study participants for some independent variables may create highly artificial circumstances. For example, it makes no sense to adjust results such that they would be valid only if there were no educational differences between young and old, in instances where such a situation would never be expected to prevail in the environmental context that provides the rationale for the research study.

Hypothesis testing research

In descriptive research we wish to discover age changes and age differences that occur in the naturalistic environment, without any intervention that has been programmed by the experimenter, and without any preformed hypotheses about the nature of the data to be described. In hypothesis-testing research, by contrast, what we want to examine the plausibiligy of a particular theory that we have formulated about development. Figure 6 gives an example of a theoretical model that specifies a particular set of progressions for adult cognitive development. This model proposes that in childhood and adolescence we acquire information. In young adulthood we apply the information and skills acquired to achieve in our job. In middle age, we assume responsibility for others, and some people few people, engage in executive behavior; that is, they are not only responsible for their family, but for large social systems. In old age, we return again to a simpler interaction with the environment and emphasize integrateion of our life-long experiences by increasingly turning inward. This is a very broad theory from which we could deduce specific experiments, to test what level of cognitive functioning is appropriate for each of these stages.

Insert Figure 6 about here

As we shall see in later sections of this chapter, there are several ways in which we can proceed to desing hypothesis-testing experiments. We may either elect to experimentally vary conditions under which the expression of a particular behavior is examined (experimental approach) or we can find instances were we can examine the relative strength of relationships between different antecedent/consequent conditions (correlational methods). In each instance we must be careful to control for possible rival interpretations that arise out of artifacts of experimental design (see later discussion of internal and external validity).

TOPIC TWO: PROBLEMS OF MEASUREMENT IN THE DEVELOPMENTAL SCIENCES

The Age Variable in Developmental Research

Age-related and time-dependent processes

We need to distinguish between some behaviors that occur only at a particular age and other behaviors that occur because the individual has been exposed to a particular environment for a certain number of years. For example, we know that people who are on the same job for too long at some point in time do not do a good job anymore (sometimes referred to as "burnout"). It has been argued that such "burnout" may be a function of increasing age. However, it is much more plausible to argue that "burnout" has nothing to do with age, burt rather that it has to do with being too long on the job. If you start a job at age 20, you might be tired of it by age 30. If you start the job at age 50 you might be tired of it age 60. This is an example of a "time-dependent" developmental change. We must be careful to distinguish such changes from developmental changes that are specifically related to age.

Directionality of time-ordered observations

Limitations and advantages for causal modeling. There is an interesting aspect about time ordered observations, and that is that time never flows backwards. What happens tomorrow cannot cause what happens today. If a researcher wishes to investigate causal relationships, it is therefore very helpful to know that what comes first must cause what comes second, not the other way around. For example, if I were to interview all my readers and find that there wass a high correlation between their number of years of schooling and their self-ratings of happyness, I could then plausibly draw the inference that it is very likely that increased years of schooling might have caused greater happiness. This inferences is much more plausible, then would the reverse interpretation, namely that because the level of happiness could have determined the number of years of schooling. On the other hand, we cannot conduct experiments in which we allow time to go forwards for one group and backwards for another. That is why we indicated earlier, that in the developmental sciences we must depend primarily upon quasi-experiments rathern than experiments.

Age as the independent or dependent variable. Most researchers think of age as an independent variable. It should be pointed out, however, that it is also possible to treat age as a dependent variable. For example, a researcher might be interested in prediciting a person's age from knowing some other individual characteristics. For example, given the amount of gray hair, one could predict from that variable what a person's age might be. In some instances such predictions may be quite useful to conceptualize age as a functional age. In other words, one might ask the question, how old a person is in terms of non-chronological indicators of aging. For example, persons who lead a healthy life may have reaction times that might be much lower than that of the average person of the same chronological age.

Time and Change: The Basic Data Matrix

Cattell's data box

We will now consider what kind of data one uses to study time and change. A useful heuristic model for this purpose is provided by the Cattell Data Cube (see Figure 7). This three-dimensional cube has three faces (often called facets). Each facet represents a two-dimensional data matrix.

Insert Figure 7 about here

Variables, persons, occasions. The most familiar facet of the Cattell Cube is its front face, which repesents a set of data collected on one measurement occasion. The columns of this matrix usually represent attributes (or variables) and the rows represent entities (individuals across which we wish to generalize our observations).

Cattell points out that the way we study individual differences most frequently is to specify a set of dependent variables, and then to study these variables across a relatively large number of entities or persons. For example, let us assume we are studying the Wechsler Adult Intelligence Test, which has 11 subtests, and we want to know what the correlation amongst those

11 tests are, then we would give the test to a large number of people. That is a typical approach which is called R-methodology. In that methodology we are interested in understanding the relationships among variables as studied across many people.

But suppose we were not really interested in the relationship among the variables, but instead we were interested in identifying different types of people who had similar patterns across the different tests. Then what we would do is to collect data where we would typically have more variables than there are people, and that would be called Q-methodology.

Thus far we have only considered a situation where each person is observed only once on each variable. We can complicate matters further by considering one of the side facets of the cube. In this case we would take a single variable and observe that variable many times in many people in what Cattell calles the T-methodology. This would be an interesting design if we wanted to know how people change over time; in fact, this is an example of a single-variable longitudinal study. In this case we can also place our focus on persons rather than on the variables, and convert T- into S-methodology if we are interested in studying types of people as they vary across occasions. Why would we want to do this? We might, for example, be interested in finding out how children who have different characteristics change over time on a particular variable.

There are two more designs we can derive from the Cattell Cube. We can take one horizontal facet; that is, study one person. This becomes a single subject design. In other words, we would take a single subject, measure that subject on a number of variables on many occasions and could study the relationship among the variables in that one individual. This P-technique approach is what one would do if one wishes tos apply statistical methods to a single subject study. Finally, we could turn this design around and convert it into an O-technique approach where we would study a single person on many variables across a few occasions. We might, for example, be interested in studying changes in brain physiology over the course of a disease. In that case, we might want to take EEG measures at many different frequencies, or locations and study these over times in order to describe the individual's electrical brain activity changing over the course of the disease.

The Cattell Data Cube is useful as an organizing principle but it should be obvious that one cannot collect data that would allow one to do all of these analyses on the same data set. The Cattell Cube illustrates at least three fundamentally different kinds of designs (see Table 1). The R and Q methods represent cross-sectional studies. The S and T methods represent longitudinal studies. The R and T methods normally involve either large numbers of individuals and a limited number of variables. The Q and S methods involve a large numbers of variables and a limited number of individuals. These O and P methods, on the other hand, represent intensive case studies on single individuals.

	Covariation Focus	Observational Entity
1. R	Variable	Person
2. Q	Person	Variable
3. P	Variable	Occasion
4. O	Occasion	Variable
5. S	Person	Occasion
6. T	Occasion	Person

Table 1. The Cattell Covariation Chart

Methods of data analysis for facets. What kind of data analyses would be appropriate for the kinds of data fitting the above model. We would normally wish to determine first of all the levels of the age related variables; what is the average level of different ages? Secondly, we would always want to know what is the variability for each level because if we do not know what the variability is, we cannot determine the amount of overlap between successive groups. Thirdly, we would typically be interested in what is called strctural analysis. We would want to know how the different variables that we have studies relate to one another. Finally, we may be interested in the analysis of the shape and rate of change.

Measurement and Quantification

Characteristics of observations

Latent constructs and observable behaviors. Perhaps the most crucial issue in the measurement and observation of behaviors of interest to the developmental psychologist is the distinction between latent constructs and observable behavior. Let us take the example of the meaurement of intelligence. In order to predict how well students will perform in college, it might be useful to have a measure of their intelligence or aptitude. However, intelligence cannot be measured directly because it is a latent construct, it is something that is unobservable. Nevertheless, there are many ways in which I can develop specific operations that may be relevant to the construct of intelligence. I could prepare an essay examination, or multiple choice examination. I could use test of mathematics, or a test of English, or many other problem solving tasks. In other words, There are many different ways in which I could measure the latent constructs, even though no single observable measure will by itself adequately represent the latent construct. What is important to note is that in psychology we are generally interested in studying the development of individuals on latent constructs. That is why we must know exactly how our observed measures relate to the latent construct of interestr. This is what we call the determination of construct validity.

Scale properties

Next we need to consider the consequences of scale properties. In the physical sciences, one measures with a meter which has a known zero, and which has equal intervals; that is, one centimeter is the same size as the next centimeter. Such a scale is usually known as a ratio scale. Psychologists rarely have similar scales available to them. The best we can usually do is to have a scale which has equal intervals, but which lacks a true zero.

Interval versus ordinal measures. In many instances, however, it may not even be possible to achieve equal interval scales, and stead we must be satisfied with using measures which only have ordinal properties. That should not disturb us too much. Many psychological issues do not require that we know precisely how much of a trait a person has. It may suffice to know whether one person has more of the trait than another.

Units of observation. It is important also to specify clearly what the units of observations are in a particular research study. In other words, are you measuring the time required to a response, or are you interested in the number of correct responses. Alternate ways in which observations can be recorded include measures of amplitude and magnitude, latency and speed, frequency and proportions of total response, and many other ways. In designing a research study, one needs to include decisions on these issues as part of developing the protocol for one's data collection.

Modeling of Change

There are additional problems for those of us interested in measuring change. Pyschological observations do not yield true scores, but also include error elements. To the extent that a score at Time 1 is not completely reliable, and that a score at Time 2 is also not entirely reliable, the difference between observations at two points in time will accumulate the errors for both occasions. In other words, the reliability of difference scores will generally be lower than the reliability of scores on either occasion. This is the reason why some people are very worried about using difference scores as their units of analysis. However, if on wishes to assess magnitude of change, it is these very scores that are of primary interest to developmental psychologists. Some psychologists would deal with the unreliability of difference scores by computing residual scores. That is, we could remove the proportion of variance attributable to the Time 1 scores from the scores obtained at Time 2. The analysis of developmental change would then be performed on the resulting residual scores.

Reliability of Measurement

One of the reasons why we are concerned about reliability is that we would like to know that a score on a test at Time 2 means the same as the score on the test at Time 1. Sometimes we simply obtain the correlation between Time 1 and Time 2, and this is what we call a test/retest correlation. Such a correlation tells us to what extent the relative position of individuals at Time 1 has been maintained at Time 2. If that correlation is high, it does not necessarily mean that people will have the same scores at Time 1 and Time 2. Table 2 gives test scores for five different persons at Time 1 and Time 2. What do you think the test/retest correlation would be for these data? It would be actually be + 1.0 because the relative position of all individuals across time is maintained, even though the absolute change is not equal across all individuals. This demonstration indicates an interesting paradox. If we have a test that has perfect test/retest reliability, it would be useless, because it would offer no help in understanding change. That is why test-retest correlation is not a good indicator of the psychometric utility of measurement operations for developmentalists. Instead we need to consider a test's internal consistency. What do we mean by internal consistency? A commonly used measure of internal consistency is the split-half reliability coefficient. This coefficient is obtained by correlating the sum of all the odd-numbered items with the sum of all the even-numbered items in a test. If that correlation were high, that would indicate high internal consistency. An alternative measure of internal consistency would be the correlation of each individual item with the total score for a test. In that instance, we take the sum of all items and correlate it with each individual item. If the average item-total correlation is high, that would be another way of showing that our test had a good internal consistency.

	Score at Time 1	Score at Time 2
Subject A	5	3
Subject B	6	7
Subject C	12	8
Subject D	14	15
Subject E	15	19

Table 2. Hypothetical scores of persons tested on two occasions

The Concept of Internal Validity

We will now consider those aspects of the concept of validity that relate to issues of importance in developmental studies. What we mean by internal validity is the basic assumption that a research study is actually testing the stated hypothesis with respect to the dependent variable of interest, rather than the influence of some third variable that the researcher is not interested in but that may lead the researcher to draw erroneous conclusions because of flaws in the desing of the study.

Campbell and Stanley (1967) have formulated a list of threats to the internal validity of quasi-experiments. It turns out that these threats are particularly relevant to developmental studies. They discussed sevent threats: History, maturation, testing (reactivity), instrumentation, statistical regression, selection, and experimental mortality.

History

Suppose we have an observation X1, and we have some kind of intervention and another observation X2. For example, suppose that we ahve measured children's vocabulary at age five and six. Let us suppose that we find no difference. Such a finding would be very surprising because we would expect that there should be gain in vocabulary at that life stage. However, there could be some occurrenced between X1 and X2 that we might label as an historical event which has affected growth of the child's vocabulary. Suppose, for example, because of some political turmoil the children we have studied could not go to school during the period intervening between our two measurements. We might erronously draw a developmental conclusion that vocabulary does not grow between age 5 and 6, one which would not necessarily bereplicated at other historical periods. If we were to repeat our study later, this event would not intervene. In that case, we would find that indeed there was a growth in vocabulary. It turns out then that history can become a threat to the validity of a longitudinal study. One way in which we could deal with that threat is by replicating the study over the same age range over another point in time.

Maturation

The second threat involves the fact that maturation proceeds while individuals are studied over time. This is a problem particularly in intervention studies. Suppose we were to have some kind of intervention between X1 and X2. The observed difference could then be due either to the intervention or due to maturation, becuase behavior may change regardless of what we do. This is why intervention studies require a non-intervention control group. Paradoxically, in developmental studies, maturation is not a threat because maturation is the very topic that we usually wish to study. In developmental studies, the real threat is that we might introduce some manipulation that would hide the effects of maturation.

Testing (Reactivity)

The next threat we need to consider is what happens when the measurement operation itself affect the subjects behavior when measured on a second occasion. Administering a test or questionnaire or introducing an observer into a behavioral situation may result in effects that might erroneously be interpreted to be maturational. On intelligence or aptitude tests, practice effects may be particularly serious over short periods of time, but long-term reactivity effects are not unknown to prevail. The reactivity threat to the validity of a longitudinal study can in some circumstances be assessed by including a control group that is tested only at T2. only.

Instrumentation

Changes in tests, experimenters, or protocol between T1 and T2 are the most obvious sources of instrumentation effects. These must be controlled for by insisting on maintaining the same procedures across all occasions of a repeated measurement studies.

Statistical Regression

The concept of statistical regression implies that when the position of an individual at T2 is predicted from that person's score at T1, the best prediction would be that the score at T2 will be closer to the mean of the group. The degree of such movement will be directly influenced by the reliability of the measurement operations; the lower the reliability the greater the regression to the mean. Of course, movement to the mean can also be attributable for differential maturation by ability level. A method known as time-reversal analysis is available to distinguish between changes due to statistical regression and developmental change.

Selection

Differential selection *per se* is not a problem in single-cohort longitudinal studies. However, it becomes the major internal validity threat in cross-sectional and other non-equivalent control group designs. It has been known for a long time that it is virtually impossible to rule out the likelihood that differences observed across groups selected at different ages may be the direct result of differential recruitment, particularly when voluntary subjects are utilized Selection effects in developmentally oriented studies have typically been considered under the general topic of cohort effects (see Topic 4 of this chapter).

Experimental Mortality (Attrition)

A major threat to the internal validity of a pretest-treatment-posttest design occurs when all subjects tested at *T1* are not available for retest at *T2*. Eperimental mortality includes subject loos due to death, disability, disappearance, or failure to cooperate for the second or ssubsequent test. There is a substantial literature that reports differences in base performance between those who appear and those who fail to appear for the second or subsequent tests. Typically the dropouts at base score lower on ability variables and describe themselves as possessing less socially desirable traits than to the retest survivors. As a consequence it has been argued that longitudinal studies may represent successively more elite groups, and after a while may no longer be sufficiently generalizable. This proposition can be tested, however, and suitable adjustments for attrition are readily available.

Threats to the External Validity of Developmental Studies

Threats to the external validity of developmental studies involve a variety of design decisions that may make it difficult or impossible to generalize the findings of a particular study to broader populations or other circumstances. Three major validity threats must be considered: Experimental units, experimental settings and measurement variables.

Experimental units

The first external validity threat is posed by the fact that a particular sample of subjects may be so unusual or unrepresentative that the findings obtained in a study with such a sample could not be replicated in any other sample. We typical deal with this issue by trying to examine random sampes of representative populations whenever we wish to describe phenomena or test propositions that are to be generalized widely. On the other hand we may have need for more limited and structured samples, when we wish examine phenomena that occur only in special populations or at particular life stages. Finally, we may want to study primarily extreme instances, when our major concern is to demonstrate that a particular relationship exists, and we are not concerned with estimating population parameters for such a relationship. Extremen instances may also be useful in the study of pathological behaviors.

Experimental settings

A second external validity problem relates to experimental settings. In order to understand basic psychological processes we often create very artificial laboratory paradigms. The question must be raised then, to what extent can we genealize from the laboratory to conditions that occur naturally in field settings. I want to caution you, however, that this distinction is sometimes very arbitrary. It may often be possible to replicate naturalistic attributes of a setting in the laboratory. On the other hand, measures taken in the field may often have very synthetic characteristics. The real issue then is to be sure that we have reasons to believe that whatever is assessed in the laboratory can indeed be translated into an application in a natural setting. Otherwise, our studies become games rather than having practical and societal significance.

Measurement Variables

The third threat to external validity relates to our measurement variables, that is, we must be sure that our measures are actually relevant to the constructs we want to measure. In developmental studies, we must be sure that there is equivalance across individuals with respect to age in cross-sectional studies. We must also have equivalence across individuals with respect to age and periods, that is the time of measurement. We must be certain that our tests or interviews have the same meaning for the groups that we are trying to compare, whether we do cross-sectional studies or compare the same individuals at different ages in longitudinal studies.

TOPIC THREE: EXPERIMENTAL MANIPULATION OF THE AGE VARIABLE

What is to be Manipulated?

We have thus far covered a number of methodological issues that apply to developmental studies in general. In the third major topic we will now zero in on ways of studying the age variable by means of experimental manipulations. We begin by asking the question what is to be manipulated.

Organismic behaviors

First of all, we may be interested in manipulating organismic behaviors. That is, behaviors which would generally emerge as a function of maturation. As developmentalists, we might be interested in assessing conditions under which the appearance of a given behavior in childhood might be accelerated, or likewise at the other end of the age range where we might want to introduce manipulations which might slow the decline of such behaviors. One substantive example involving this type of manipulation would be a simple reaction time study. As we know, reaction time decreases in childhood and incrases as we get older. We might be interested in modeling what the conditions would be that might either accelerate or slow these changes. Other examples might be biofeedback sstudies as well as studies of temperament. All of these are thought to be maturationally determined, but are subject to environemental manipulation in terms of their rate and time of occurrence.

Acquired behaviors

A second class of variables where we might study age effects experimentally are behaviors which ar typically acquired. For example, we might manipulate age effects with respect to cognitive abilities. We might test a proposition that it might be possible to affect the acquisition of crystallized abilities, but not necessarily of fluid abilities. Crystallized abilities are culturally specific and must be acquired by individuals in any given culture; these are what we term the pragmatics of intelligence . Fluid abilities are those intellectual processes which presumably are characteristic of maturational development, and which might be thought of as the mechanics of intelligence. Another class of acquired behavior that may be experimentally manipulated is memory. Again we can test propositions above the conditions under which memory might show age differences.

Response styles

A third class of variables that we may want to bring under experimental control with respect to age might be studies of response styles. To give you an example of response styles, I might mention the topic learned helplessness. Learned helplessness occurs in situations over which one has no control. The classical case is an animal experiment where an animal is shocked in a cage, and there was no way for the animal to escape. After a while the animal simply sits and takes the punishments, and that is what we called learned helplessness. We know that there are age differences in the extent to which people show this phenomenon. One could design experiments to try to find out what is the nature of these age differences. Another example of a response style that is interesting from the standpoint of development is what is called flexibility/rigidity. What we mean by flexibility is that a person is able to respond appropriately even though the situation has changes, rather than to continue to respond in a way which was appropriate in another situation but which is not appropriate in the new situation. We know that when we compare young and old people, the older people tend to be less flexible. Again we could design experiments to check out what might be the reasons for the difference.

Ethical issues

There are some ethical issues involved when one conducts experiments with people. In theory, the most clear cut experiment requires that you are able to show that you can change a behavior both in a negative and positive direction. But in studies with humans, it is obviously not ethical to change their behavior to something that is worse than what they are currently experiencing. For example, it is reasonable to study whether you can accelerate the appearance of positive cognitive behavior in children, but most people would find it morally unacceptable to design experiments which would retard the development of children.

Age-Comparative Studies

Simple young/old design

What kind of experimental studies on age can then be suggested? First of all, we have a class of studies that are generally called age comparative studies. The most simple design is a young-old design, taking a young group and an old group and giving them some measures and see whether they differ. As we have already pointed out that is not a very useful design because it suffers from all of the threats to internal validity that we have discussed earlier.

Age by treatment interaction design

A better design is what we might call an age by treatment interaction design. In this case we would measure a young and an old group of subjects on a dependent variable, say performance on an associative memory test. Suppose we would then give both groups 10 practice trials. In this case we would probably find that the old group did less well than the young group. It is not surprising that the young group at the base line would perform higher than the old group. We know that it would almost be impossible to match them so that they would perform at the same level. But if it was true that the reason for the age difference could be attributed to the fact that the old individuals had had less practice on this task, we would then expect that the old group would benefit more from practice than the young group. The study I have just described is really a simple analysis of variance design where we have a two by two analysis of variance with one between group factor and one within group factor. What is important to note is that the two main effects are really not too interesting because if we pick two samples that are sufficiently far apart in age we will always get an age effect. And if our experimental manipulation has been properly chosen then we should get a practice effect. The first result is to be expected and the second is a necessary condition for our study. What is really of interest for a developmentalist is the interaction because if that interaction is not significant then our manipulation does not explain the age differences. If it is significant, then the intervention may indeed explain part of the age difference. There are problems as well with this design, but nevertheless it is the design is quite common in the experimental aging literature, particularly in fields such as learning and memory. What the reader should remember when reading this literature is to focus their attention upon the crucial interaction effect.

Molar-molecular decomposition design

A recent study by Salthouse demonstrated older typists were able to type as well as younger typists. As would be expected, their simple reaction time was longer than that of younger typists, but they were apparently able to compensate. This study uses a relatively new but powerful design for the study of the aging called molar-molecular decomposition. In this research design two groups of people are aidentified that are different in age but are equivalent in performance. It is then possible to study other relevant characteristics of the two groups that differ which may help us to understand why it is possible some times for older persons to perform as well as younger persons. This approach can help us understand what are some of the compensating mechanisms that make it possible for older people to function as well or better than younger persons in many practical situations. This is an important area of research because if we have a better underderstaning of what these compensating mechanisms are we might be able to teach people how to compensate for some of the losses that occur with aging.

Within Age Group Studies.

A second class of experimental aging studies are what we might call within age group studies. I mentioned earlier that such studies require some theoretical models. First of all we must specify some assumptions about the nature of the age change. For example, are we talking about irreversible decrement or some other model and what are some of the assumptions about the measurement model? By that I mean what are the measures or observations that will be used to operationalize the construct that you are interested in.

Design of intervention studies

Most within group studies typically involve some kind of intervention. There are basically two different designs for intervention studies. The first is a post-test only design which depends upon effective random assignment of subjects to the treatment and control groups. The second, more desirable and most justifiable design is what is called a pretest-posttest design. That is, there is a pretest followed by an intervention which in turn is follwed by a posttest. In this design there also is a control group which does not receive the intervention. This design equates for and thus controls for the effects of practice or reactivity across the two groups. Unless one has very large samples, however, one can not be certain that there has beenn effective random assignment of subjects to the intervention and control groups. This is why the pretest/posttest design is a relatively conservative approach, because it permits comparison of the change from pre- to posttest in both treatment and control group, rather then relying entirely upon a post-test only comparison.

Control groups

There are several different kinds of control groups. One kind of control group is simply one which takes the pretest and posttest but no intervention. This is sometimes called a no-contact control group. Some people would argue that had an experimenter simply spent time with the subjects an effect would occur. This is tested by using a contact control group which does not receive the intervention but something else happens in lieu of the intervention that is not relevent to the experiment but that creates an interaction with the experimenter. For example, in our previously mentioned practice study we would have given the experimental group practice. The no-contact control group could have had no treatment, but the contact control group might have had a conversation about psychological testing.

Transfer of training issues

We also need to be concerned about transfer of training issues. Some psychologists would argue that the effects of any kinds of intervention may be very general. Simply paying attention to your subject may produce a postive result. The reader may be familiar with the term "Hawthorne Effect". Industrial psychologists have found that if you make a change in a factory, no matter what the change you make, there will be a temporary positive effect. This is why we need to have in our pre-posttest design both a measure of the dependent variable that is of primary interest to us, as well as some other dependent variable for which we would predict that our intervention should produce no change. If we obtain an effect for the variable for which we predicted that there should not be an effect then we have encountered transfer of training and we do not know whether our intervention is specific to the desired dependent variable.

Example: The Seattle Training Study

We were interested in determining whether it would be possible to improve performance of older people who had either shown decline or had remained stable on the psychological abilities of spatial orientation and inductive reasoning. This kind of study requires that some previous information is available. In our case we studied people on whom we had prior data over a 14 year period.

Classification procedures

The first question was how do we classify subjects who had remained stable or had declined? This is of course an arbitrary decision. But you have to have some criterion and the criterion we used was to specify the confidence interval of one standard error of measurement (SEM) about the individual's base score. If subjects in 1984 were within one SEM of their score in 1970, we classified them as stable. If they fell more than one SEM below their 1970 score, then we classified them as having declined.

Time-reversal techniques to assess regression effects

Earlier I discussed the problem of statistical regression as a threat to the validity of a study. The threat was investigated by dividing our subjects into three groups: a middle group, a high group and a low group. Then we computed the scores for these groups in 1970, and at pretest in 1984. Figure 8 shows the data for the three groups for all our subjects and for those subjects only who had declined significantly (left side of graphs). At first glance, it appears that there may have been regression effects, since the average score for the top group declines. However, in the presence of gross linear regression effects, we would also expect that the average score for the lowest group would go up, but it does not. Next we classify our subject into the top, bottom and middle third in terms of their scores in 1984. If there was statistical regression operating, we would now expect the 1970 average for the top group to be lower and the bottom group to be higher (right side of graphs). However, the actual data shows the lines are quite parallel. Our data conssequently represent true developmental change and not regression.

Insert Figure 8 about here

Designing the intervention procedures

Next, we designed intervention techniques to see if we could improve the performance of our subjects on the two abilities we were studying. The way this is done is to conduct a task analysis, and then develop exercises that are relevant to the ability. It is important to be certain that we do not teach the subjects how to take the test. Rather we teach them skills which are relevant for performing well on the test. For example, in spatial orientation, we started out by giving subjects pictures of hands in different locations. We had subjects first practice by rotating cards and then rotating the figures in their heads. We also gave them map exercises and other exercises relevant to developing spatial orientation concepts. We developed similar strategies for inductive reasoning.

Qualitative vs quantitative change

As mentioned earlier, we are not interested in teaching people how to do better on a particular test. What we want to accomplish is to help them to improve on a construct (basic ability). To assess the construct, we had to develop multiple measures for each construct. For example, on inductive reasoning, we had four different tests and the same approach was used for spatial orientation. Recall now our earlier comments on transfer of training. We wanted to be certain that our subjects improved on those abilities that we had targeted for training but not on others irrelevant abilities. That is why we also included in our battery tests of the constructs of perceptual speed, of numerical ability, and of verbal ability. The hypothesis we were testing was that those subjects exposed to space training would improve only on that ability but not on the others, and those subjects assigned to the inductive reasoning would improve on that ability. Figure 9 presents results on the average change between pre- and posttest on five abilities. As can be seen, those subjects who reasoning training improved signficantly more on the reasoning ability as compared with the group who did not. The group that was trained on space similarly improved signficantly on the space ability. This design confirmed our hypothesis that we were able to do ability-specific training. The study furthermore is an example of a design using two different experimental treatments, where each experimental group is used as the control group for the other treatment. In addition to the group results, we also paid close attention to the proportion of individuals who were trained successfully. Approximately 40% of those subjects who had declined were brought back to their performance level in 1970, before they declined. This study consequently provided an important demonstration that much of what we think of as age-related decline may be attributable to disuse, lack of practice and experience.

Insert Figure 9 about here

TOPIC FOUR: QUASI-EXPERIMENTAL DESIGNS

LONGITUDINAL VS. CROSS-SECTIONAL METHODOLOGY

The Age-Cohort-Period Model

In the last topic in this chapter we will consider more formally the distinction between cross-sectional and longitudinal methods, and related issues in the design of studies in developmental psychology. For the last 20 years or so, we have expressed this problem in terms of what we call the "general developmental model".

The basic model: R = f(A,C,T)

We can express a change in behavior as a function of three components: Age, cohort and period. The first component is the chronological age (A) of an individual at the time we observe that individual. The second component is what we call cohort (C) membership. By that term we mean those individuals who have entered the environment at the same time; for example, everybody born in 1970. There are also other definitions for cohort. For example, everybody who entered college in 1985, regardless of age, may be thought of as belonging to a cohort. Or we could define a cohort as everybody who was employed in a factory when it first opened. Traditionally, however, cohort is defined in terms of date of birth. The third component is the period or time of observation (T), the time a particular measurement was taken.

Alternative definitions for chronlogical age

There are possible alternative difinitions for age. For example, we could substitute for chronological age some other concept of age. We could substitute a biological definition, say measured by cardiac output, or metabolic rates. We could also have some sociological definitions in terms of particular life stage. In other words, we could assign a person's age relative to the age when they got married. Or we could assign the same social age to everyone who had their first child in the same year. All of these would be social ages rather than chronological ages. We could alternatively assign psychological ages. The most faimilar of these would be the concept of mental age. The point I am trying to make is that the use of chronological age in developmental research is quite arbitrary. Indeed, there is some current interest in substituting more meaningful age dimensions. The major advantage of calendar age is that we all know what we are talking about.

Traditional Single Level Research Strategies Derived from the General Model

In terms of the above considerations, we will first examine traditional single level research strategies that can be derived from the general model. We will consider here the standard cross-sectional and single-cohort longitudinal methods, as well as the less well-known time lage method.

Cross-sectional studies

When we consider an age difference found in a cross-sectional study, we may note that such difference will be composed of the term A plus the term C (as defined above). T will not be involved because a cross-sectional study is conducted at one point in time, so time is a constant. In the case of a single cross-sectional study, we do not know whether an observed age difference is due to a maturational change or whether it represents different life experiences between the two groups that are compared. We have already discussed the selection problems involved. The major advantage of a cross-sectional study is that it can be done relatively quickly. Since only one measurement is taken from each individual, one does not have to worry about the threats of reactivity, instrumentation, experimental mortality or history. But the problem still remains, that we do not know whether an observed difference involves age or cohort effects. In child studies, if you compare age groups that are close to each other, the cohort effect is probably quite small. It may not be unreasonable in that case to think that cross-sectional age differences may provide a good estimate of age changes. But in studies of adults, where age ranges are large and developmental change is slow, it is very likely that the cohort component may be larger in magnitude than the age component.

Figure 10 gives an example of this problem. The top part of this figure shows the median income in the U. S. of workers according to different ages, as obtained from a cross-sectional study. It is apparent thate the highest income was reached in the decade of the 40s. From this data, we might conclude that workers' income rose until midlife and declined sharply thereafter. The bottom par t of the figure provides projections of income for three group of workers in successive cohrots. These projections suggest that workers' income actually rose slightly throughout life, but that the older persons started at a lower income level when they were young, because over time workers' income has gone up. When you do a cross-sectional study, it looks as if income is going down with increasing age because the different cohorts are measured at one point in time confounding the workers' entry income by cohort with changes in income over time. We may conclude from this demonstration that in the presence of possitive cohort effects, cross-sectional studies will make it appear as if there are negative age changes. On the other hand, if cohort effects are negative, then cross-sectional data will make it appear as if people are getting better as they get older, even though this is not true.

Insert Figure 10 about here

Longitudinal studies

Let us next concern ourselves with longitudinal studies. If the same people are followed over time, cohort is constant, but we do not know if an observed change is attributable to maturation or to some transient historical effect. That is the single-cohort longitudinal study can not distinguish between maturational change and societal changes or treatment interventions that occurred over the time period of our study. This may not be a problem in the study of psysiological development, but it certainly poses limitations on behavioral science inquiries. Longitudinal studies do not suffer from selection problems since the same people are followed. However, longitudinal studies do suffer from the validity threats of attrition, instrumentation, statistical regressiona, and reactivity that we discussed earlier. The major advantage of a longitudinal study is that it permits the identification of individual patterns of change. For example, it would be impossible to study differences between early and late maturing children by means of a cross-sectional study. Nor would it be possible to investigate individual differences in rates and timing of age-related decline without a longitudinal study. The major disadvantage of a longitudinal study is that it takes time. And if you are dealing with phenomena that occur very slowly, it may take a long time.

Time-lag studies

There is a third type of study of interest to developmental and social psychologists which may be unfamiliar to the reader, is the time-lag study. In a time-lag study, we compare individuals of a given age with individuals of the same age at a different point of time. For example, if we wanted to examine the performance of students entering college over time, that would be an example of a time lag study. On average, students enter college at the same age. We might be interested to know, for example, whether the scores on entrance examinations were higher in 1987 than in 1985. A time-lag difference is also confounded because it has cohort and time components. However, age is constant because we are comparing two groups of people at the same age. The difference in performance between two college classes in 1985 and 1987 could be due either to differences in the population of entering students or could be due to some historical events that occurred between 1985 and 1987.

Sequential Multi-level Strategies

We noted that the single level strategies always hold one of the components of the general developmental model constant, while confounding the other two components. When we have a problem that has more unknowns than equations, the thing to do is to increase the number of equations by means of replication. What one can do then is introduce sequential designs. These designs are essentially replications of the simple designs. There are two ways in which this can be done: longitudinal and cross-sectional sequences. Figure 11 shows a schematic from which we can identify all of the various sequential design combinations available from the general developmental model.

Insert Figure 11 about here

Data collection strategies

The two boxed bottom rows in Figure 11 show an example of a longitudinal sequence. In this case we have followed two cohorts (born in 1950 and 1960 respectively) from age 10 to age 70. The two boxed diagonal rows represent what is known as a cross-sectional sequence. What we have done here is to conduct two cross-sectional studies, one in 1960 and one in 1970. In each of these years we collected data for all of the ages of interest (10 to 70 years). Note that the longitudinal data could come from the same subjects measured repeatedly, or from independent samples of the same cohort taken sequentially. The cross-sectional data, of course, must come from different individuals for each age.

Analysis strategies

No single study would obviously contain all the data in our schematic. Indeed, for purposes of analysis we would typically take some subsets having certain characteristics. From the sequential data matrix we can derive three sequential designs that will allow us to separate two of the three components of developmental change while confounding the third. For example, we can take a subset of two groups of people born in 1880 and 1890 whom we had tested at ages 10 and 20. The first group would have been tested in 1890 and 1900, the second group in 1900 and 1910.

When we compare the two groups at the same age, we can get an average value for age. We can also get an average value for the the two cohorts across the same age range. This is a cohort-sequential design which allows us to disentangle the *A* and *C* effects while confounding *T*. We have another design that allows us to break apart age and time of measurement. In a time-sequential design we measure at two points in time for two or more ages so that we can differentiate between age and time of measurement. The example in Figure 11 is a data set where we have groups that are 10 and 20 years old tested in 1920 and 1930. These groups are drawn from cohorts born ina 1900, 1910, and 1920. In this design we disentangle the *A* and *T* effects while confounding *C*. Finally we can break apart cohort and time in what is called a cross-sequential design. The example in our schematic would we the data set involving cohorts born in 1880 and 1890 which are tested in 1920 and 1930 respectively. The older cohort would be 40 and 50 years of age, while the younger cohort would be 30 and 40 years of age, respectively. This design disentangles the *C* and *T* effects while confounding *A*.

Schaie's "most efficient design"

The sequential designs at first glance seem very complicated. There is a way, though that might simplify them for the interested reader. Figure 12 shows what has been called "Schaie's most efficient design". We begin our research by doing a cross-sectional study at one point in time over the age range in which we are interested. This range is from age 30 to 60 in our hypothetical study in which we include four groups of individuals. There is very little we can do with the initial data collection other than the usual statements made from cross-sectional studies. A great many additional options become available, however, if we can follow these individuals at a second point in time. In our example, after 10 years, what we have are four short-term longitudinal studies. After the first assessement we had information only on the age difference between the different groups, say between ages 30 and 40. The short-term longitudinal studies allow us direct estimates of age changes over four ten year age periods; in other words, we have in 10 years actually obtained longitudinal data over a 40-year span. At the second data collection point we also collect data on new samples that are of equivalent age to our retested subjects, as well as another sample representing the age level of our youngest group at base test. We can now assess the effects of reactivity (practice) by comparing individuals of the same age at first and second test. Adding the additional samples will also permit us to conduct time-sequential and cross-sequential analyses. Moreover, we can contrast base scores in our longitudinal sample for those individuals whom we retested compared to those who dropped out to conduct experimental mortality analyses. A third data point would include repeated testing of the <u>T1</u> and <u>T2</u> subjects would allow replication of the longitudinal findings, extending them for another decade, and permit cohort-sequential analyses.

Insert Figure 12 about here

<u>Example: The Seattle Longitudinal Study</u>

I shall now give you an empirical example of a complex sequential study. Figure 13 present the basic design of what is known as the Seattle Longitudinal Study.. That study began in 1956 as my doctoral dissertation. I was interested in studying age changes in adult intelligence. In 1956 we assessed 500 persons, equally distributed by age from 25 to 70 with an equal number of men and women. In 1963, we retested as many of these people as we could, around 300 people. But we also assessed another 1,000 persons over the age range from 22 to 77 who were a random sample from the same orginal population. We continued this study in 1970 and were able to retest about 170 of the original subjects and 500 of the second group, and we again assessed another 700 new subjects from age 72 to 84. This is exactly the kind of sampling plan discussed earlier. We repeated this approach again in 1977 and 1984, but I do not want to make this too complicated, so I will only talk about the first three waves of the study.

Insert Figure 13 about here

Let us now consider some of the findings presented in Figure 14 for the Space test from Thurstone's primary mental abilities measure. Consider first the left upper quadrant of this figure. The solid lines represent cross-sectional data, the solid circles for 1956 and the open circles for 1963. The dashed lines are longitudinal data, representing change in the same people from 1956 to 1963. As you can see until about age 60, the longitudinal data are basically flat. Now look at the data in the lower left quadrant for the people we first picked up in 1963. The solid cireclesare the scores in 1963 and the open circles those for 1970. Next, let us look at the three cross-sectional studies, 1956, 1963 and 1970, represented in the upper right quadrant. Note that successive groups keep getting better and better on the space test, the most recent cross-section data are always on top of the early data. These differences most likely occur as a function of changes in levels of education and increasingly favorable living conditions. I suspect these cohort differences would even be greater in China because your society is experiencing faster changes.

Insert Figure 14 about here

Finally, let us examine the lower left quadrant of Figure 14. This figure shows the longitudinal data for those individuals who were assessed three times. These are age changes within individuals. Again I want you to note that for the young adult group the trends are actually up, for the middle aged perfectly flat, and in advanced age there is substantial decline. These data also provided some interesting information about choosing cohort boundaries. I arbitrarily picked mine in seven year segments. Our data, however, informs us that there are probably only three cohort groupings. It is possible to identify historical events which separate these three groups. The oldest people received their education before the first World War, the next two groups received their education between World War I and World War II, and the younger groups received their education after World War II. Both of these wars produced very marked changes in American society. Sommetimes psychological data can indeed reflect historical events.

I would now like to call attention to some other data from the Seattle Longitudinal Study. The five measures that we have consistently assessed in this study, include Verbal Meaning, which is a measure of recognition vocabulary, Spatial Orientation, Inductive Reasoning, Number skills and Word Fluency, which is a measure of vocabulary recall. The data in Figure 15 are the result of an experimental mortality analysis, comparing average performance by birth cohort at base test for those individuals who did or did not return to be tested. The solid dots are always the subjects whot came back, and the open dots are those who did not. These data suggest that at virtually all ages dropout is more frequent for the less able, but there are exceptions for any findings like this. Notice that for Number and Word fluency, there is a reversal for the youngest group. This information probably means that there was a greater mobility in young people who had skills who went into jobs that require these skills. This group would be more likely to drop out because they moved to another place for employment reasons.

Insert Figure 15 about here

SOME FINAL COMMENTS

Before concluding this chapter I would like to reiterate some important points on the interpretation of age-related findings in psychology. Much of the work that psychologists do is concerned with making statements about differences between groups defined in one way or in another. Depending on what characteristics they have, we often make a big deal out of showing that a difference between two groups is statistically significant. Let me remind you that a finding of statistical significance only means that the particular finding is non-chance and has a high probability of being replicated in another study. For example, if a repored group differenceis significant at the 1% probability level, this means that if the same study were repeatred 100 times, we should get the same results 99 times out of 100. This finding does not necessarily mean that the magnitude of the observed difference is very large. There are two ways in which we can make judgements about the importance of a group difference in terms of magnitude. One is to ask how large is the difference or change in relation to the variability around the mean in the total population. We do this by dividing the mean difference by the population standard deviation for the variable of interest. Generally we would expect a difference to be of practical importance if it attains a magnitude of at least one-half of a standard deviation. If it is smaller than that, it may be a viable interesting phenomenon, but it does not have much practical significance because of the enormous overlap between two groups you are comparing. Table 3 shows the magnitude of change from initial level in the early 20s to advanced age on several abilities. In the case of Verbal Meaning, for example, there is virtually no change from young adulthood to middle age. By the time the young-old stage is reached in the 60s are, there is a drop of something like one-third of a standard of deviation, which is a very small drop. Only by the time of advanced old age, at about 80 years, is there a change which is more than three-fourths of a standard deviation. On verbal behavior, until young-old age the annual average rate of loss is as small as 1/100 of a standard deviation. Even though the observed changes are very reliable, the actual magnitude and the rates at which they occur is very small.

Insert Table 3 about here

Another way of looking at the issue of overlap is to create a scatter plot of your observations. Figure 16 shows a plot of 1629 individuals on Verbal Meaning. The correlation with age is about 0.5, which suggests that approximately 25% of individual differences can be explained by age, but 75% can not! Note that there are many older people who perform less well than most younger people, but there are some even at age 88 who did better than did some 25 year olds. And certainly, at least until age 80, there is virtually complete overlap. Even at age 80 there are some people who do much better than average young people. We must, consequently, be very careful before we attribute differences between young and old adults as being due to age, when gender differences may be due to many other causes.

Insert Figure 16 about here

There are many other issues that remain to be covered, but perhaps I have given you enough food for thought and I do not want you to suffer from indigestion. The references attached to this chapter will direct you to more extensive discussions of the matters discussed in this chapter to and other methodological issues of interest to developmental psychologists. My workshop audience was a very responsive one, and I much enjoyed my first opportunity to interact with such an enthusiastic group of Chinese graduate students.being with you. I hope the readers of this chapter will also be stimulated by these remarks and trust that these summary comments will entice their curiosity enough to continue to deepen their familiarity with the topics here discussed.

BIBLIOGRAPHY

Major References:

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). Life-span developmental psychology: Introduction to research methods. Hillsdale, NJ: Erlbaum. (All Topics).

Birren, J. E., & Schaie, K. W. (Eds.). (1977). Handbook of the psychology of aging. New York: Van Nostrand Reinhold. (Topics 1 & 4).

Birren, J. E., & Schaie, K. W. (Eds.). (1985). Handbook of the psychology of aging (2nd ed.). New York: Van Nostrand Reinhold. (Topics 1 & 2).

Kausler, D. H. (1982). Experimental psychology and human aging. New York: Wiley. (Topic 3).

Nesselroade, J. R., & Baltes, P. B. (Eds.). (1979). Longitudinal research in the study of behavior and aging. New York: Academic Press. (Topic 3).

Schaie, K. W. (Ed.). (1983). Longitudinal studies of adult psychological development. New York: Guilford Press. (Topics 2 & 4).

Schaie, K. W., Campbell, R. T., Meredith, W., & Rawlings, S. C. (Eds.). (1988). Methodological issues in aging research. New York: Springer Publishing Co. (All Topics).

Schaie, K. W., & Willis, S. L. (1986). Adult development and aging. Boston: Little, Brown & Co. (Topics 1 & 4).

Wolman. B. B. (1982). Handbook of developmental psychology. Englewood Cliffs, NJ: Prentice-Hall. (Topics 2 & 4).

Additional References:

Krauss, I. K. (1980). Between- and within-group comparisons in aging research. In L. W. Poon (Ed.), Aging in the 1980s (pp. 542-551). Washington: American Psychological Association. (Topic 3).

Salthouse, T. A. (1987). The role of experience in cognitive aging. In K. W. Schaie (Ed.), Annual review of gerontology and geriatrics (Vol. 7, pp. 135-158). New York: Springer Publishing Co. (Topic 3).

Schaie, K. W. (1983). The Seattle Longitudinal Study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), Longitudinal studies of adult psychological development (pp. 64-135). New York: Guilford Press. (Topic 4).

Schaie, K. W. (1986). Beyond calendar definitions of age, cohort and period: The general developmental model revisited. Developmental Review, 7, 252-277. (Topics 2 & 4).

Schaie, K. W. (1987). Age changes in perceptual speed: A seven-year longitudinal study. Paper presented at the Cognition and Aging Conference, Atlanta, GA. (Topics 2 and 4).

Schaie, K. W. (1988). Internal validity threats in studies of adult cognitive development. In M. L. Howe & C. J. Brainerd (Eds.), Cognitive development in adulthood: Progress in cognitive development research. New York: Springer Verlag. (Topics 2 and 4).

Schaie, K. W. (1988). Variability in cognitive function in the elderly: Implications for social participation. In A. Woodhead, M. A. Bender & R. C. Leonard (Eds.), Phenotypic variation in populations: Relevance to risk assessment (pp. 191-212). New York: Plenum. (Topics 2 and 4).

Schaie, K. W., Orshowsky, S. J., & Parham, I. A. (1982). Measuring age and socio-cultural change: The case of race and life satisfaction. In R. C. Manuel (Ed.), Minority aging: Sociological and social psychological issues (pp. 223-230). Westport, CT: Greenwood Press. (Topic 4).

Schaie, K. W., & Parham, I. A. (1974). Social responsibility in adulthood: Ontogenetic and socio-cultural changes. Journal of Personality and Social Psychology, 30, 483-492. (Topics 2 & 4).

Schaie, K. W., & Willis, S. L. (1986). Can decline in adult intellectual functioning be reversed? Developmental Psychology, 22, 223-232. (Topic 3).

Willis, S. L. (1987). Cognitive training and everyday competence. In K. W. Schaie (Ed.), Annual review of gerontology and geriatrics (Vol. 7, pp. 159-188). New York: Springer Publishing Co. (Topic 3).

Willis, S. L., & Schaie, K. W. (1986). Training the elderly on the ability factors of spatial orientation and inductive reasoning. Psychology and Aging, 1, 239-247. (Topic 3).

FIGURE CAPTIONS

Figure 1. Projected population (in millions) by age and sex: United States, 1985-2050. Source: U.S. Bureau of the Census. (1982). Projections of the population of the United States. Current Population Report, Series P-25, No. 922. Washington: U.S. Government Printing Office.

Figure 2. Average years of life remaining at various ages. Source: National of Health Statistics. (1983). Changing mortality patterns, health services utilization and health care expenditures: United States, 1978-2003. Vital and Health Statistics, Series 3, No. 23.

Figure 3. Number of psychological aging publications per year from 1880 to 1979. Source: Poon, L. W., & Welford, A. T. (1980). Prologue: an historical perspective. In L. W. Poon (Ed.), Aging in the 1980's (p. xiv). Washington: American Psychological Association.

Figure 4. The interaction of three systems of influences regulating the nature of life span development. Source: Baltes, P. B. (1979). Life-span developmental psychology: Some converging observations on history and theory. In P. B. Baltes & O. G. Brim, Jr. (Eds.), Life-span development and behavior (Vol.

2). New York: Academic Press.

Figure 5. Alternate models of aging: Irreversible decrement, decrement with compensation, and adult stability. Source: Author.

Figure 6. Stages of adult cognitive development. Source: Schaie, K. W. (1977/78). Toward a stage theory of adult cognitive devleopment. Journal of Aging and Human Development, 8, 129-138.

Figure 7. The Cattell data cube. Source: Baltes, P. B., Nesselroade, J. R., & Reese, H. W. (1977). Life-span developmental psychology: Introduction to research methods. Monterey, CA: Brooks/Cole.

Figure 8. An example of the application of time-reversal technique for the study of regression effects in longitudinal data. Source: Schaie, K. W., & Willis, S. L. (1986). Can decline in adult cognitive functioning be reversed? Developmental Psychology, 22, 223-232.

Figure 9. An example of transfer of training in an intervention study. Source: Willis, S. L., & Schaie, K. W. (1986). Training the elderly on the ability factors of spatial orientation and inductive reasoning. Psychology and Aging, 1,

239-247.

Figure 10. Medican income of white American males by age: An example of the "cross-sectional" fallacy. Source: Schaie, K. W., & Willis, S. L. (1986). Adult development and aging (p. 26). Boston: Little, Brown & Co.

Figure 11. Schematic showing cross-sectional and longitudinal sequences and the modesl of adnalysis deduced from the developmental model. Source: Schaie, K. W. (1983). What can we learn from the longitudinal study of adult psychological development. In K. W. Schaie (Ed.), Longitudinal studies of adult psychological development (p. 9). New York: Guilford Press.

Figure 12. Schaie's "most efficient design": A data collection scheme for sequential analyses. Source: Schaie, K. W., & Willis, S. L. (1986). Adult development and aging (p. 30). Boston: Little, Brown & Co.

Figure 13. Design of the Seattle Longitudinal Study. Source: Author.

Figure 14. Cross=sectional and longitudinal findings for the Space test. Source: Schaie, K. W., & Labouvie-Vier, G. (1974). Generational versus ontogenetic components of change in adult

cognitive behavior. Developmental Psychology, 10, 305-320.

Figure 15. An example of experimental mortality analysis. Source: Schaie, K. W., Labouvie-Vief, G., & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. Journal of Gerontology, 28, 328-334.

Figure 16. Scatter plot by age for the Verbal Meaning test. Source: Schaie, K. W. (1988). Variability in cognitive function in the elderly: Implications for social participation. In A. Woodhead, M. A. Bender & R. C. Leonard (Eds.), Phenotypic variation in populations: Relevance to risk assessment (pp. 191-212). New York: Plenum.

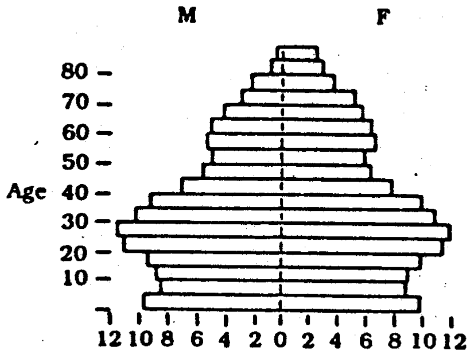
Table 3

Magnitude of Age Changes in Standard Deviation Units by Ability and Cohort

	From Initial Level			From Peak Level		
	Verbal Meaning	Spatial Orient.	Induct. Reason.	Verbal Meaning	Spatial Orient.	Induct. Reason.
Middle-aged	+ .07	- .15	- .07	- .20	- .30	- .23
Young-Old	- .30	- .29	- .39	- .41	- .29	- .48
Old-Old	- .73	- .74	- .96	- .73	- .92	- .96

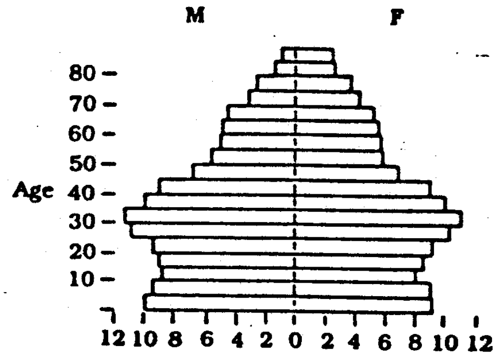
Annual Rate of Change

	Verbal Meaning	Spatial Orient.	Induct. Reason.
Middle-Aged	+ .002	- .003	- .005
Young-Old	- .010	- .011	- .012
Old-Old	- .035	- .026	- .033



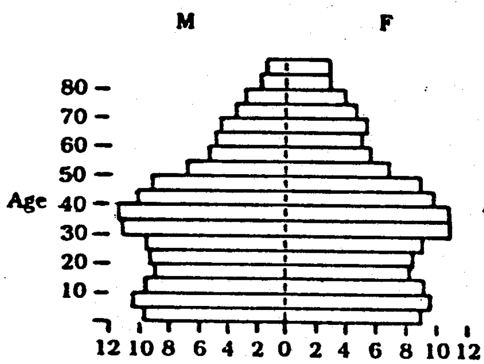
1985

Median = 31.5



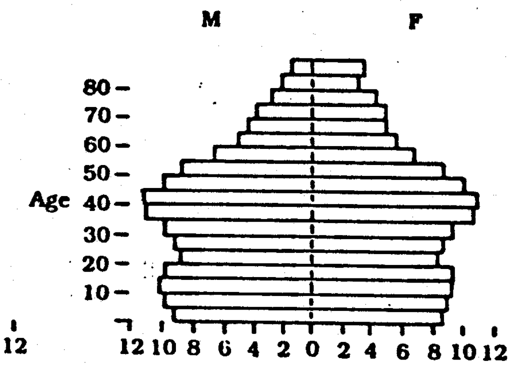
1990

Median = 33.0



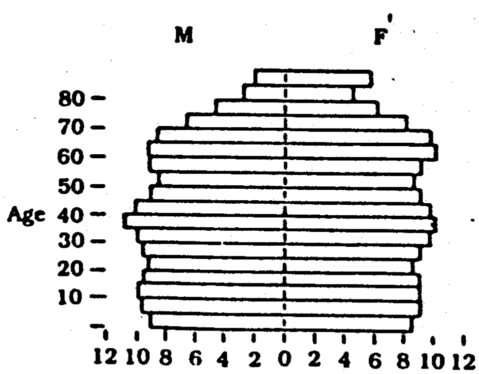
1995

Median = 34.7



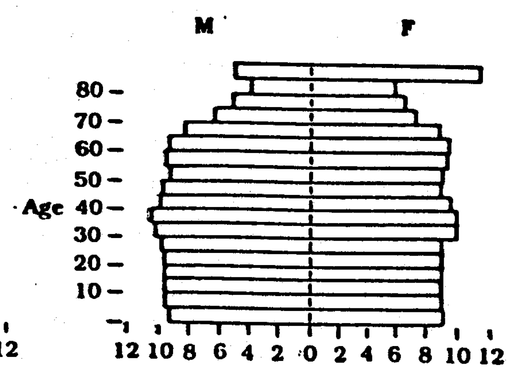
2000

Median = 36.3



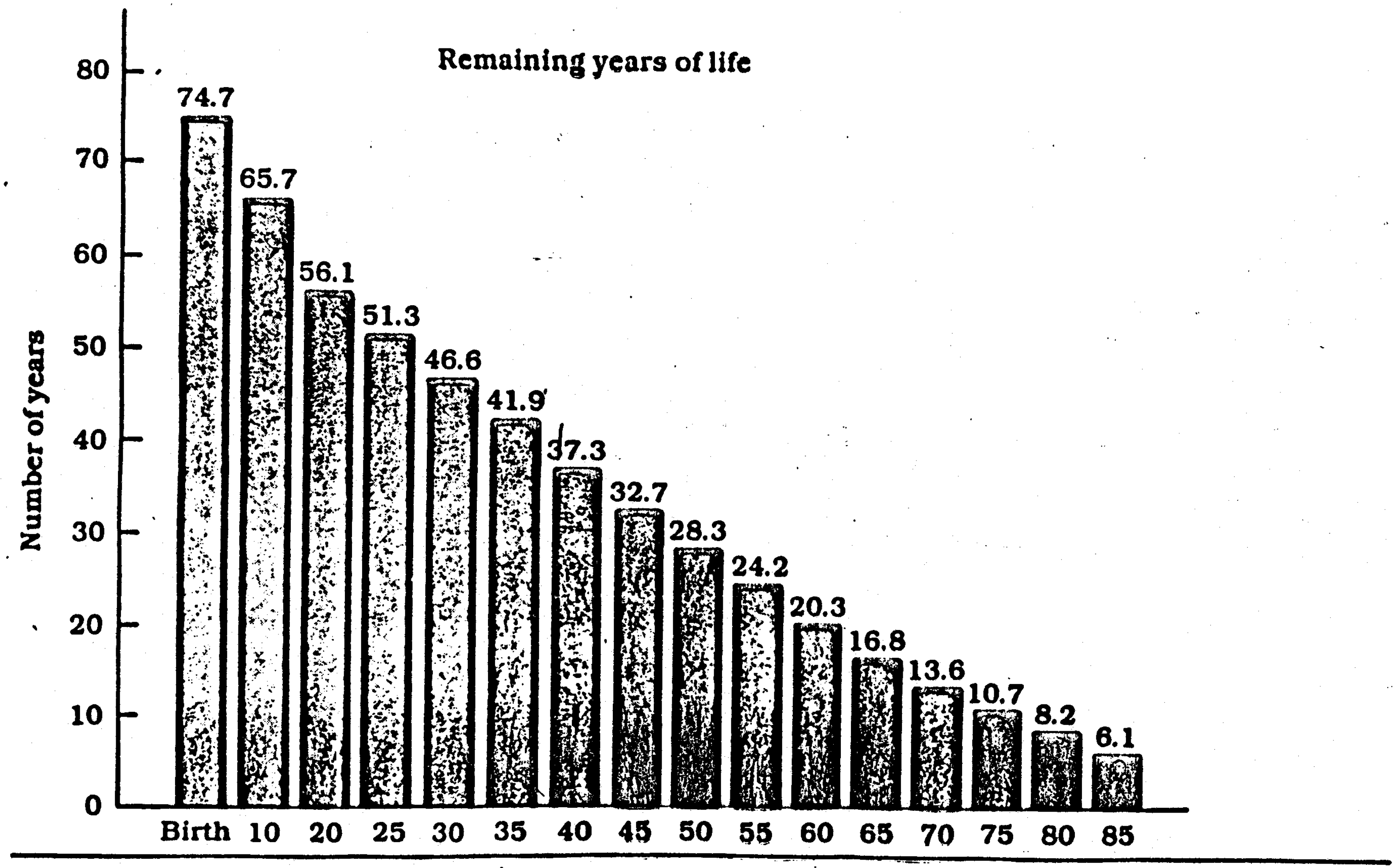
2025

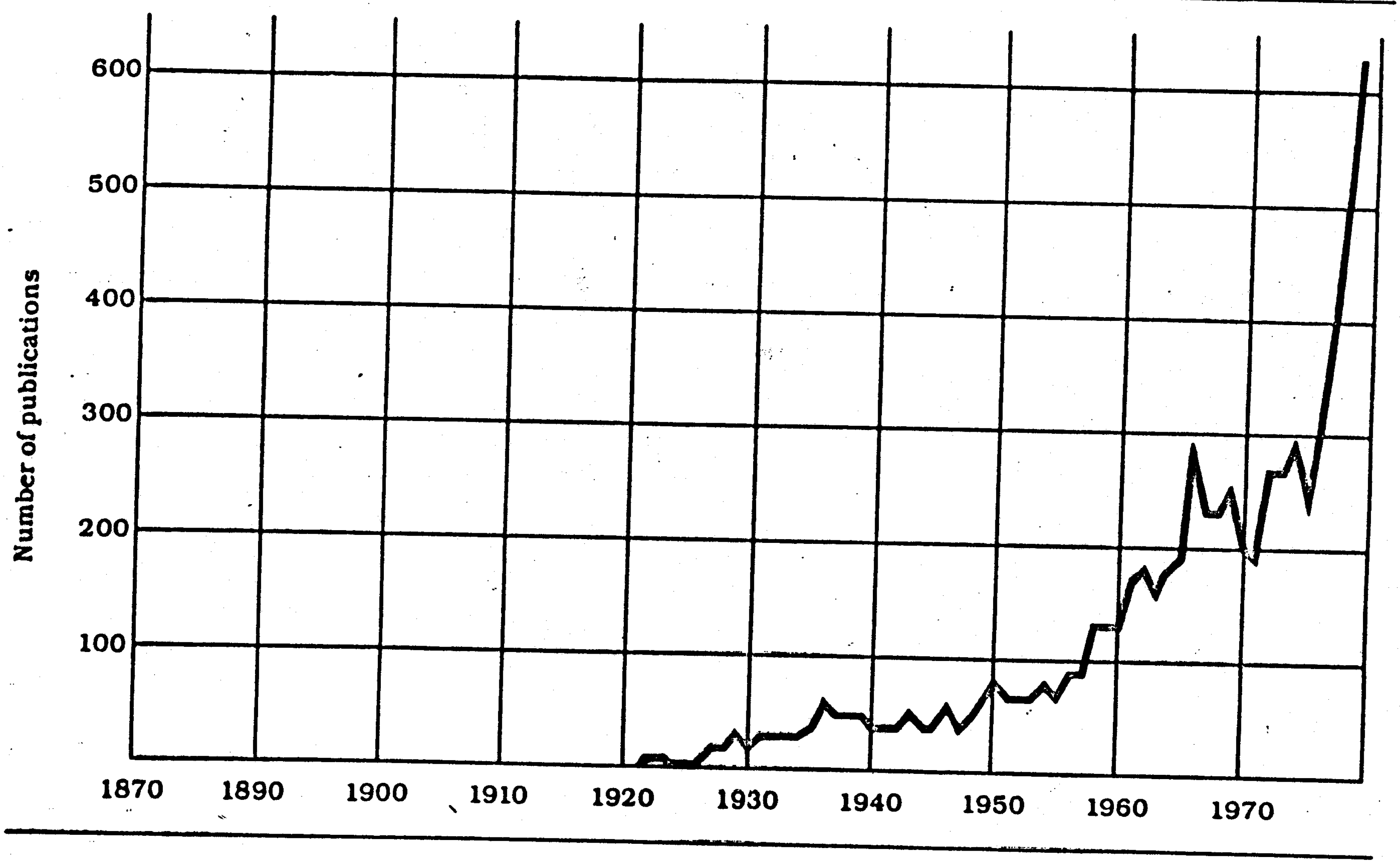
Median = 40.0

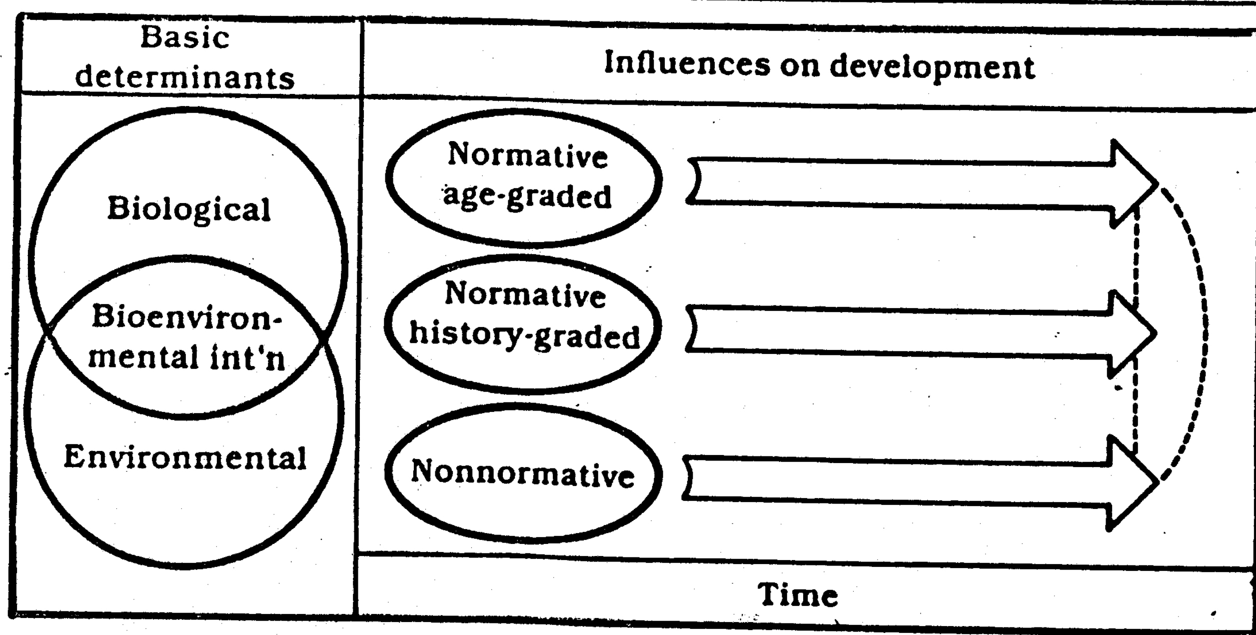


2050

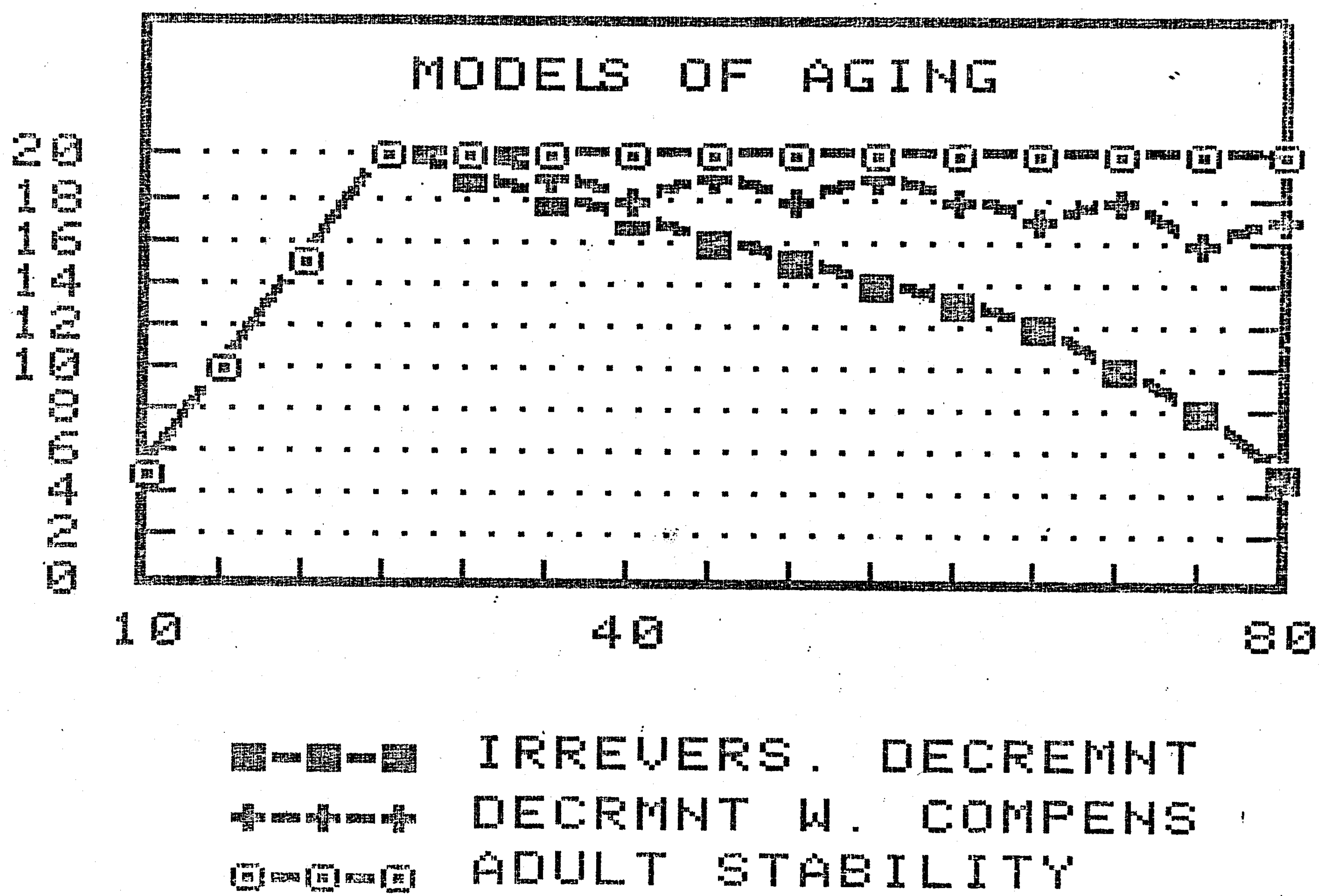
Median = 41.6

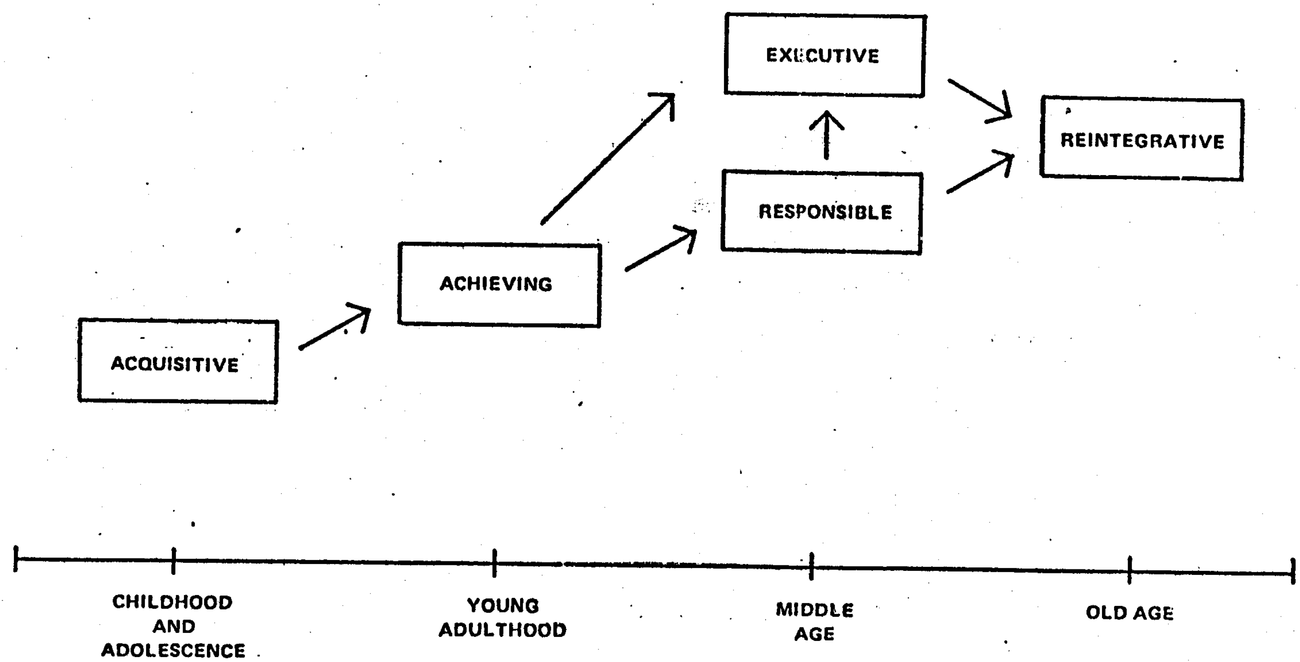

Remaining years of life
Number of years
80
70
60
50
40
30
20
10
0
74.7
65.7
56.1
51.3
46.6
41.9
37.3
32.7
28.3
24.2
20.3
16.8
13.6
10.7
8.2
6.1
Birth
10
20
25
30
35
40
45
50
55
60
65
70
75
80
85

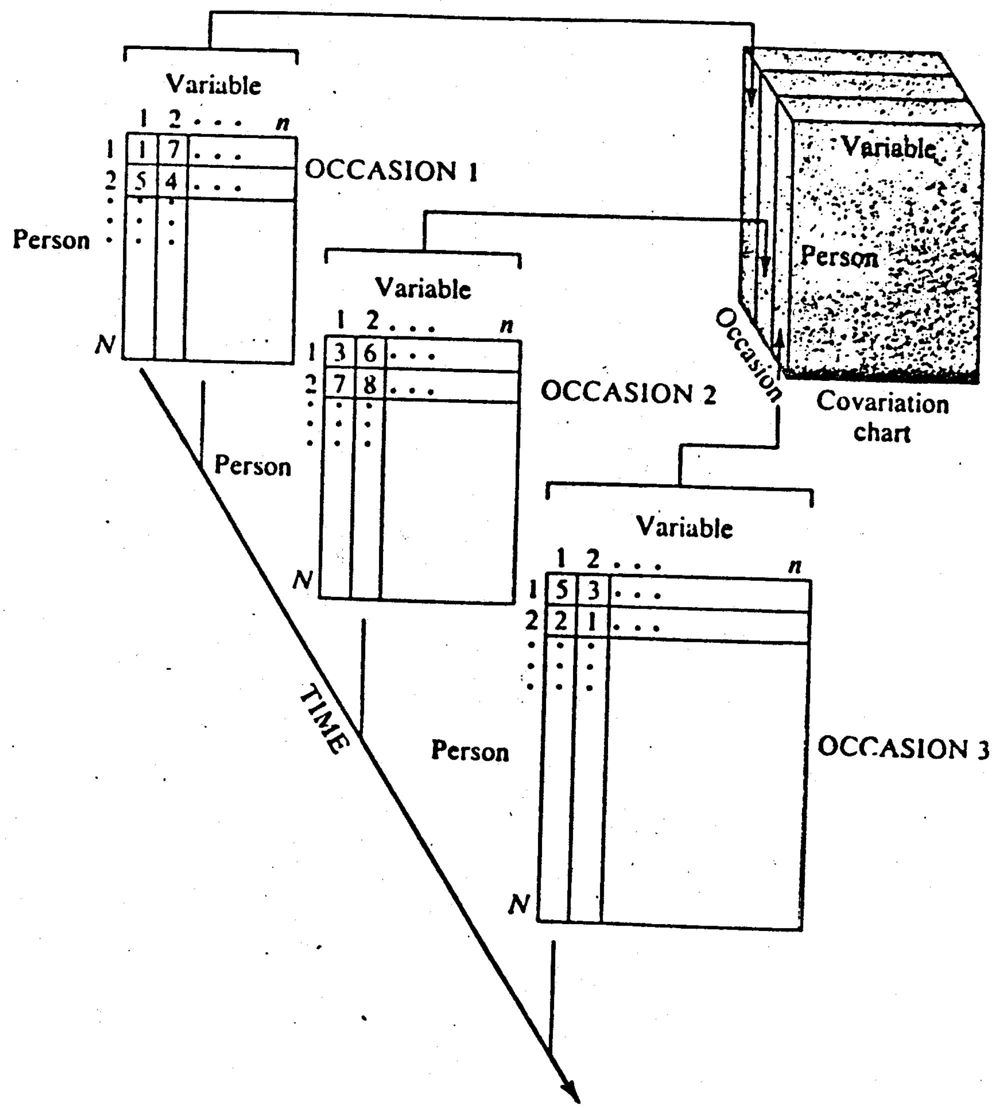

Number of publications
600
500
400
300
200
100
1870
1890
1900
1910
1920
1930
1940
1950
1960
1970



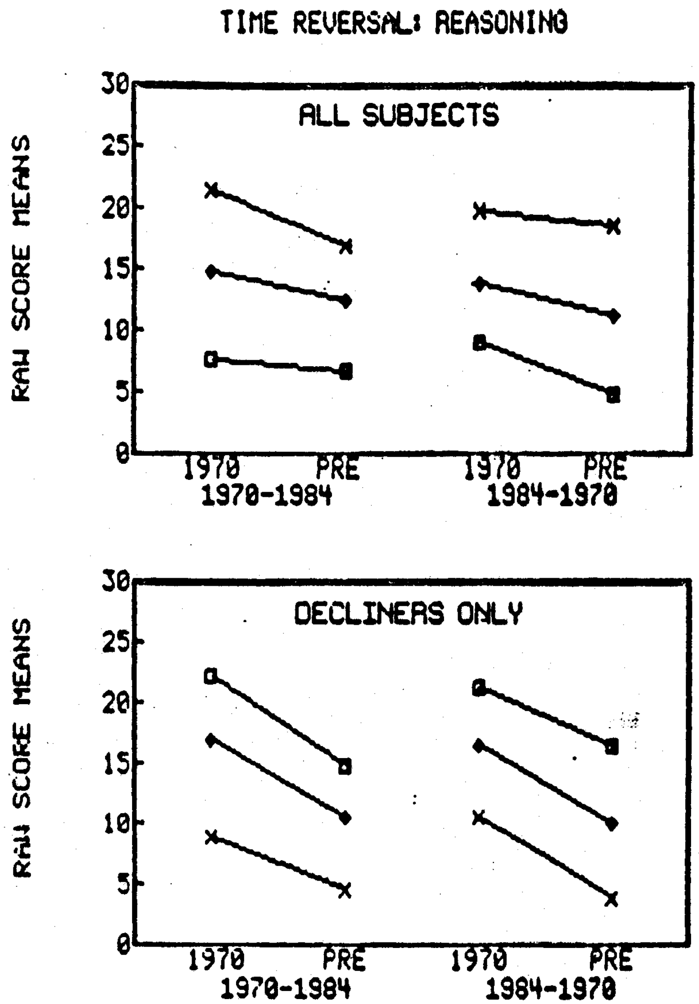
Basic determinants
Influences on development
Biological
Bioenviron-mental int'n
Environmental
Normative age-graded
Normative history-graded
Nonnormative
Time

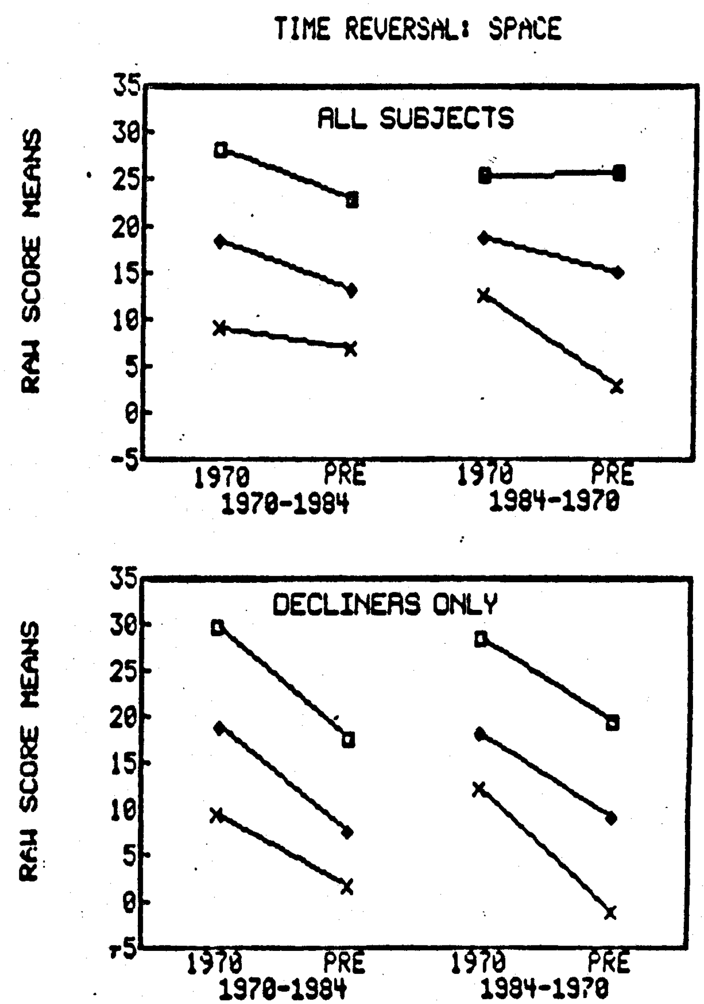

MODELS OF AGING
0
2
4
6
8
10
12
14
16
18
20
10
40
80
IRREVERS. DECREMNT
DECRMNT W. COMPENS
ADULT STABILITY


EXECUTIVE
REINTEGRATIVE
RESPONSIBLE
ACHIEVING
ACQUISITIVE
CHILDHOOD AND ADOLESCENCE
YOUNG ADULTHOOD
MIDDLE AGE
OLD AGE

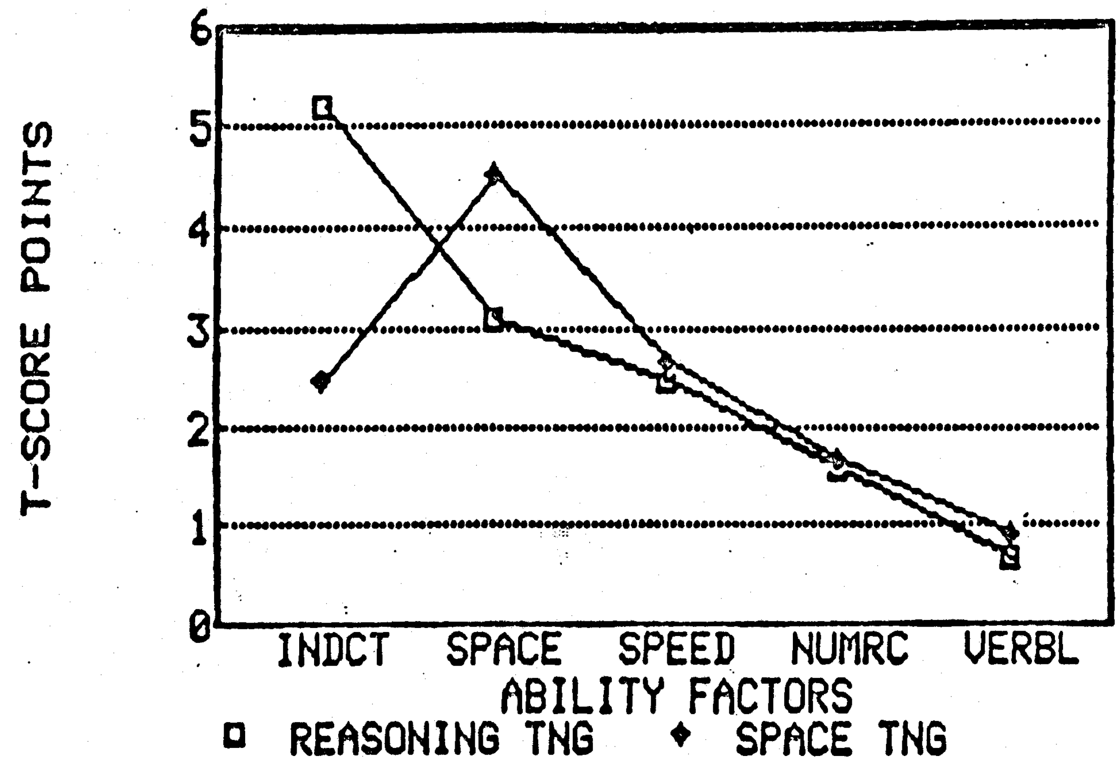


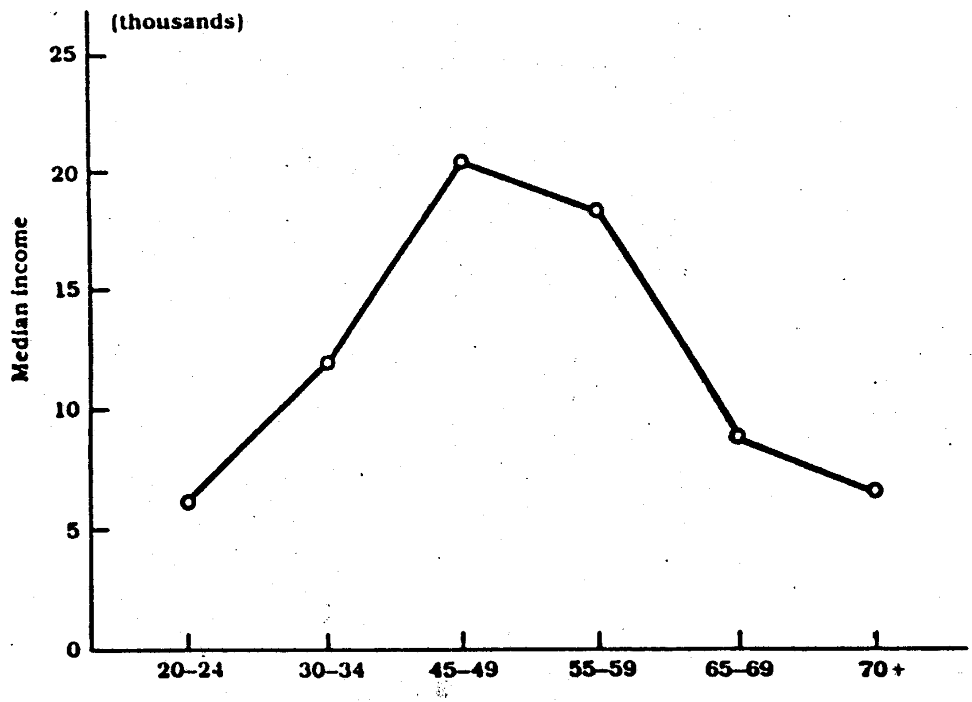
Variable
1 2 . . . n
1 1 7 . . .
2 5 4 . . .
Person
N
OCCASION 1
Variable
1 2 . . . n
1 3 6 . . .
2 7 8 . . .
Person
N
OCCASION 2
Variable
1 2 . . . n
1 5 3 . . .
2 2 1 . . .
Person
N
OCCASION 3
TIME
Variable
Person
Occasion
Covariation chart



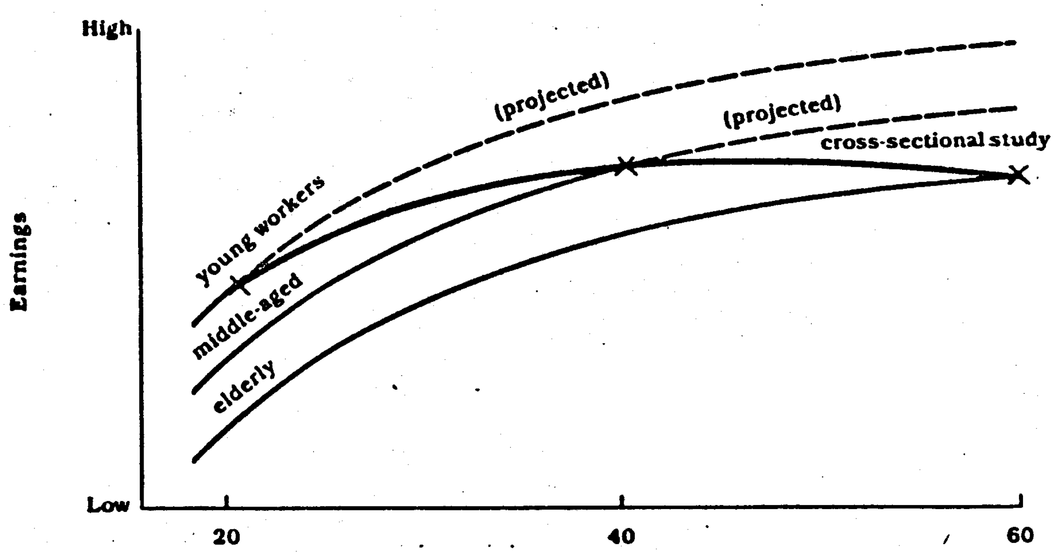


□ LOWEST THIRD ◆ MIDDLE THIRD × HIGHEST THIRD

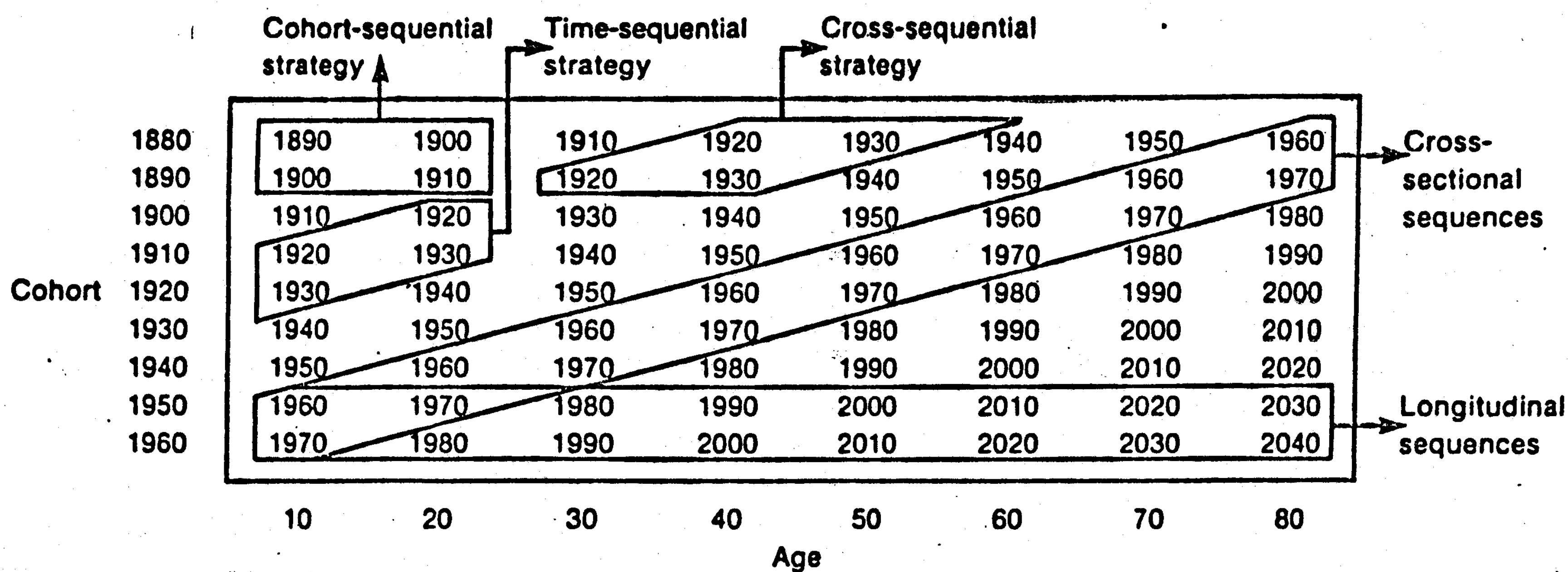

T-SCORE POINTS
0
1
2
3
4
5
6
INDCT
SPACE
SPEED
NUMRC
VERBL
ABILITY FACTORS
REASONING TNG
SPACE TNG

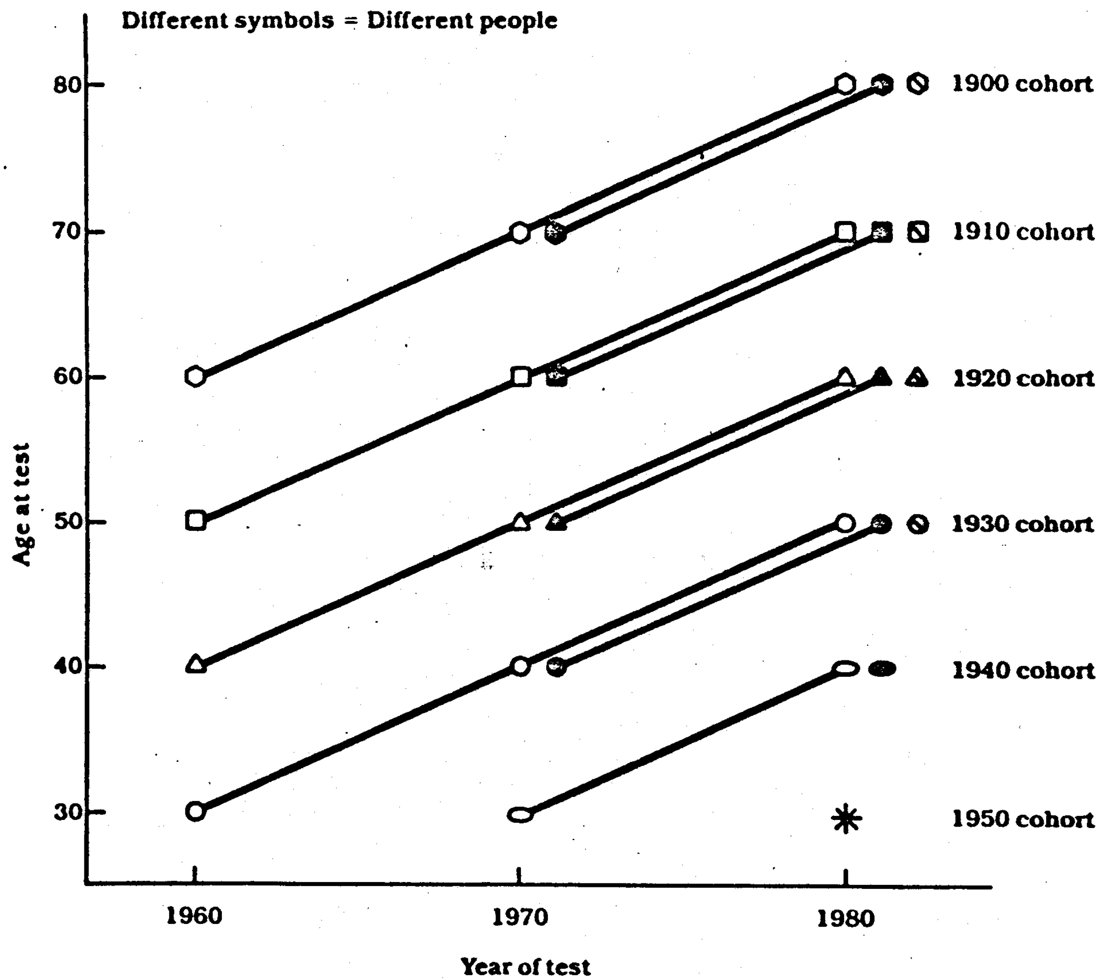

(thousands)
25
20
15
10
5
0
Median income
20–24
30–34
45–49
55–59
65–69
70 +

a) Age


High
Low
Earnings
(projected)
(projected)
cross-sectional study
young workers
middle-aged
elderly
20
40
60

b) Age


Cohort-sequential strategy
Time-sequential strategy
Cross-sequential strategy
Cohort
1880 1890 1900 1910 1920 1930 1940 1950 1960
1890 1900 1910 1920 1930 1940 1950 1960 1970
1900 1910 1920 1930 1940 1950 1960 1970 1980
1910 1920 1930 1940 1950 1960 1970 1980 1990
1920 1930 1940 1950 1960 1970 1980 1990 2000
1930 1940 1950 1960 1970 1980 1990 2000 2010
1940 1950 1960 1970 1980 1990 2000 2010 2020
1950 1960 1970 1980 1990 2000 2010 2020 2030
1960 1970 1980 1990 2000 2010 2020 2030 2040
10 20 30 40 50 60 70 80
Age
Cross-sectional sequences
Longitudinal sequences



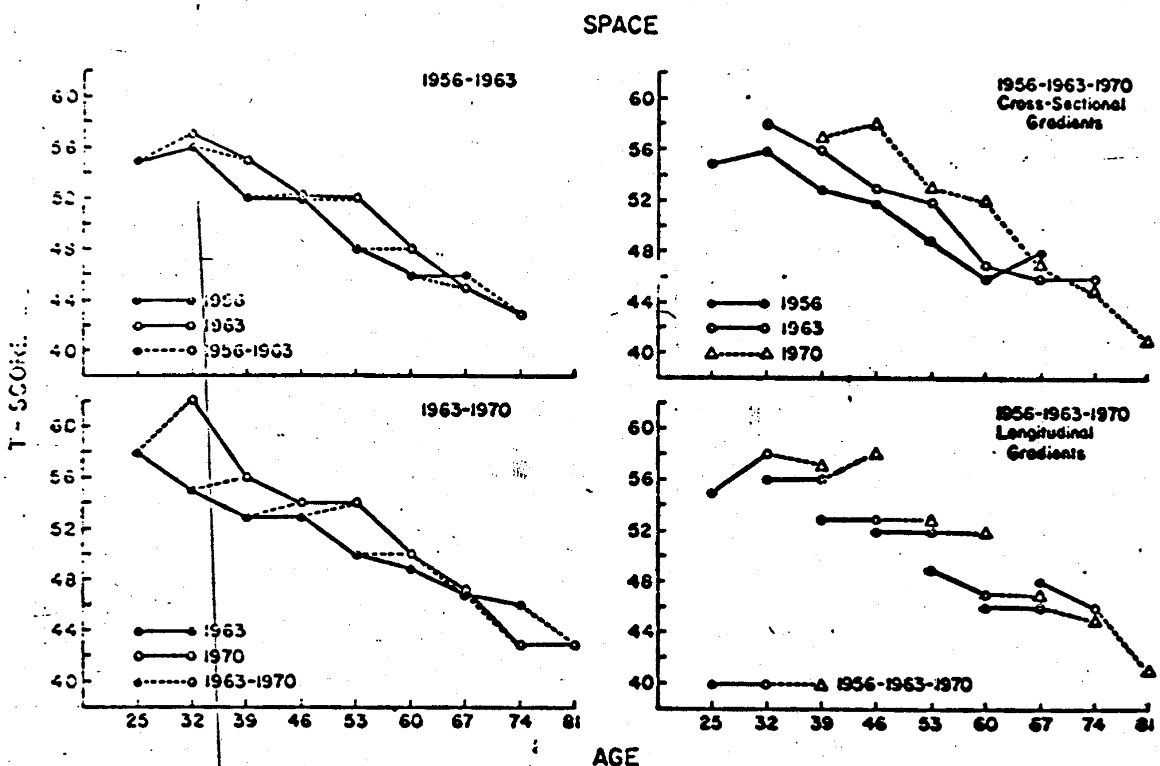
Different symbols = Different people
80
70
60
50
40
30
Age at test
1900 cohort
1910 cohort
1920 cohort
1930 cohort
1940 cohort
1950 cohort
1960
1970
1980
Year of test

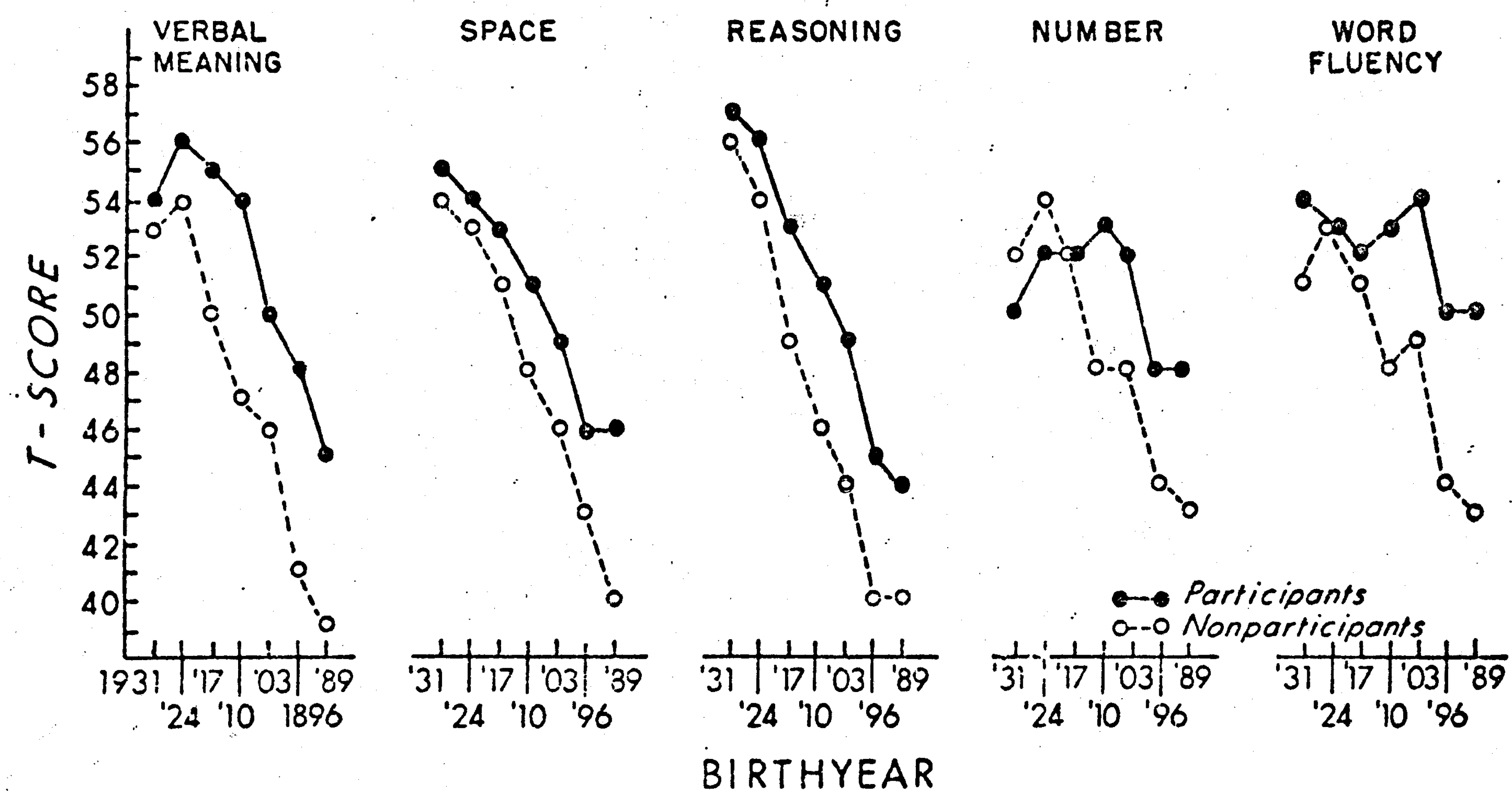
Design of the

SEATTLE LONGTUDINAL STUDY

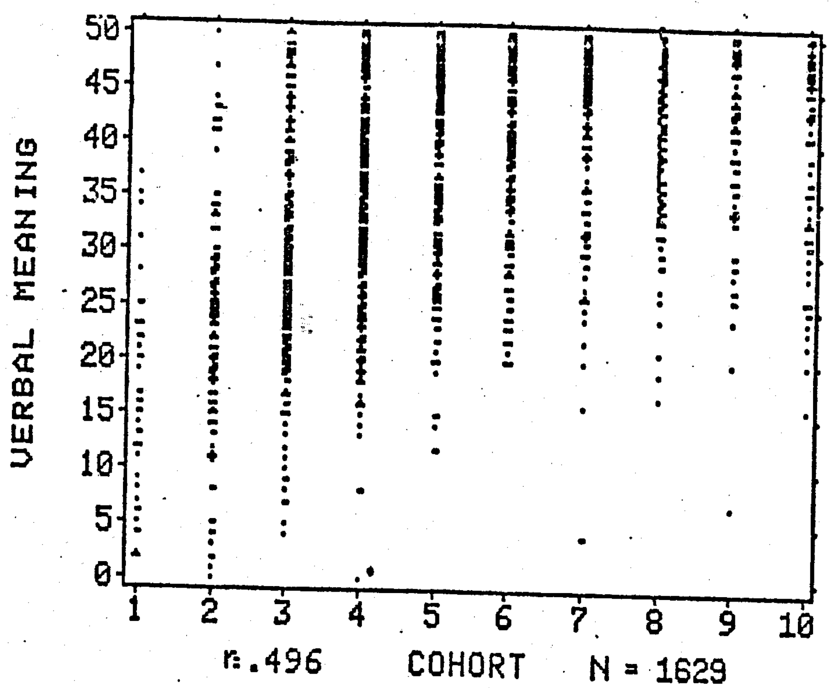
Study Waves

1956	1963	1970	1977	1984
S1T1 (N=500)	S1T2 (N=303)	S1T3 (N=162)	S1T4 (N=130)	S1T5 (N= 97)
	S2T1 (N=997)	S2T2 (N=420)	S2T3 (N=337)	S2T4 (N=204)
		S3T1 (N=705)	S3T2 (N=340)	S3T3 (N=225)
			S4T1 (N=612)	S4T2 (N=294)
				S5T1 (N=628)


SPACE
1956-1963
1956
1963
1956-1963
1963-1970
1963
1970
1963-1970
1956-1963-1970
Cross-Sectional
Gradients
1956
1963
1970
1956-1963-1970
Longitudinal
Gradients
1956-1963-1970
T - SCORE
AGE
25
32
39
46
53
60
67
74
81
40
44
48
52
56
60



VERBAL MEANING
SPACE
REASONING
NUMBER
WORD FLUENCY
T - SCORE
58
56
54
52
50
48
46
44
42
40
1931 '24 '17 '10 '03 1896 '89
'31 '24 '17 '10 '03 '96 '39
'31 '24 '17 '10 '03 '96 '89
'31 '24 '17 '10 '03 '96 '89
'31 '24 '17 '10 '03 '96 '89
BIRTHYEAR
Participants
Nonparticipants


VERBAL MEANING
0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7
8
9
10
r=.496
COHORT
N = 1629