


1972

A review of "longitudinal study" in developmental psychology

Emily H. Finley
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

 Part of the [Longitudinal Data Analysis and Time Series Commons](#), and the [Psychology Commons](#)
Let us know how access to this document benefits you.

Recommended Citation

Finley, Emily H., "A review of "longitudinal study" in developmental psychology" (1972). *Dissertations and Theses*. Paper 3043.

<https://doi.org/10.15760/etd.2856>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

AN ABSTRACT OF THE THESIS OF Emily H. Finley for the Master of Science in Psychology presented July 28, 1972.

Title: A Review of "Longitudinal Study" in Developmental Psychology

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

[REDACTED]
Frank Wesley, Chairman

[REDACTED]
Vincent Glaudin

[REDACTED]
Max Réed

The purpose of this library research thesis is to review the "longitudinal study" in terms of problems and present use. A preliminary search of the literature on longitudinal method revealed problems centering around two areas: (1) definition of "longitudinal study" and (2) practical problems of method itself. The purpose of this thesis then is to explore through a search of books and journals the following questions:

1. How can "longitudinal study" be defined?
2. What problems are inherent in the study of the same individuals over time and how can these problems be solved?

A third question which emerges from these two is:

3. How is "longitudinal study" being used today?

This thesis differentiates traditional longitudinal study from other methods of study: the cross-sectional study, the time-lag study, the experimental study, the retrospective study, and the study from records. Each of these methods of study is reviewed according to its unique problems and best uses and compared with the longitudinal study. Finally, the traditional longitudinal study is defined as the study: (1) of individual change under natural conditions not controlled by the experimenter, (2) which proceeds over time from the present to the future by measuring the same individuals repeatedly, and (3) which retains individuality of data in analyses.

Some problem areas of longitudinal study are delineated which are either unique to this method or especially difficult. The following problems related to planning the study are reviewed: definition of study objectives, selection of method of study, statistical methods, cost, post hoc analysis and replication of the study, time factor in longitudinal study, and the problem of allowing variables to operate freely. Cultural shift and attrition are especially emphasized. The dilemma is examined which is posed by sample selection with its related problems of randomization and generalizability of the study, together with the problems of repeated measurements and selection of control groups. These problems are illustrated with studies from the literature.

Not only are these problems delineated but considerable evidence is shown that we have already started to accumulate data that will permit their solution. This paper presents a number of studies which have considered these problems separately or as a side issue of a

study on some other topic. Some recommendations for further research in problem areas are suggested.

At the same time that this thesis notes differentiation of the longitudinal study from other studies, it also notes integration of results of longitudinal studies with results of other studies. The tenet adopted here is: scientific knowledge is cumulative and not dependent on one crucial experiment.

Trends in recent longitudinal studies are found to be toward more strict observance of scientific protocols and toward limitation of time and objectives of the study. When objectives of the study are well defined and time is limited to only enough for specified change to take place, many of the problems of longitudinal study are reduced to manageable proportions.

Although modern studies are of improved quality, longitudinal method is not being sufficiently used today to supply the demand for this type of data. Longitudinal study is necessary to answer some of the questions in developmental psychology. We have no alternative but to continue to develop this important research tool.

A REVIEW OF "LONGITUDINAL STUDY"
IN DEVELOPMENTAL PSYCHOLOGY

by
Emily H. Finley


A thesis submitted in partial fulfillment of the
requirements for the degree of

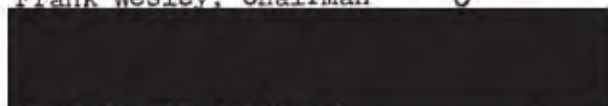
MASTER OF SCIENCE
in
PSYCHOLOGY

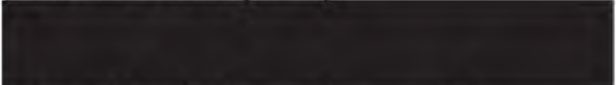
Portland State University
1972

TO THE OFFICE OF GRADUATE STUDIES:

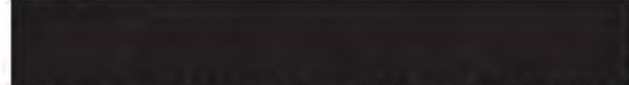
The members of the Committee approve the thesis of
Emily H. Finley presented July 28, 1972

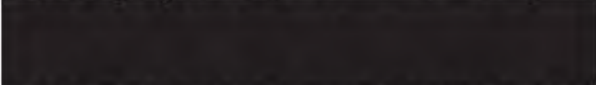

Frank Wesley, Chairman


Vincent Glaudin


Max Reed

APPROVED:


Robert E. Jones, Jr., Head, Department of Psychology


David T. Clark, Dean of Graduate Studies

July 28, 1972

ACKNOWLEDGMENTS

I thank the professors who gave their time to serve on my committee, Frank Wesley as chairman, Vincent Glaudin, and Max Reed. They have given me the freedom to pursue the topic of this thesis where the literature indicated without bias within the limitations of my own judgement. On the other hand their reading of the manuscript and commentary have been of valuable assistance. I especially thank them for the confidence they have shown in me.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	vii
 CHAPTER	
I INTRODUCTION	1
Purpose of Thesis	1
History of Longitudinal Study	1
Need for Review of Longitudinal Study	3
II "LONGITUDINAL STUDY" DEFINED	7
Relationship of "Longitudinal Study" to Cross-sectional Study	8
Cross-sectional, Time-lag and Longitudinal Studies Individuality of Data Retained in Longitudinal Study Summary of Cross-sectional Study	
Relationship between Longitudinal and Experimental Studies	14
Thorndike's Study Longitudinal Study as Study of "Things are They Are Naturally" Comparison of Experimental and Longitudinal Design Interpretation of "Causality" in Experimental and Longitudinal Studies Predictive Usefulness of Longitudinal Study Summary of Relationship between Longitudinal and Experimental Studies	

CHAPTER	PAGE
Retrospective and Prospective Studies	19
Yarrow, Campbell, and Burton Study Summary of Prospective and Retrospective Study	
Study from Past Records	21
Advantages of Study from Records Problems of Study from Records Summary of Study from Records	
Traditional Longitudinal Study	25
"Longitudinal Study" Defined "What's in a Name?" A Traditional Longitudinal Study Summary of Traditional Longitudinal Study	
III PROBLEMS OF LONGITUDINAL STUDY	29
Planning the Study	29
Definition of Study Objective Selection of Method of Study Correlation Analysis Cost Time Factor in Longitudinal Study <u>Post Hoc</u> Analysis and Replication of the Study Problem of Allowing Variables to Operate Freely Summary of Planning the Study	
Cultural Shift	43
Age-Time of Testing Confound Age-Time of Birth Confound Strategies Cultural Shift in Developmental Psychology Itself Summary of Cultural Shift	
Sample Selection, Randomization and Generalizability of the Study	50
The Basic Dilemma Selection of Control Groups Problem of Repeated Measurements Miscellaneous Problems of Sampling	

CHAPTER	PAGE
Summary of Problems Related to Sample Selection	
Attrition	55
Attrition a Soluble Problem	
Bias of the Sample due to Attrition	
Bias when Attrition is not Controlled	
When Sample Size is Reduced by Experimenter	
Mobility and Attrition	
Summary of Attrition	
IV REVIEW OF RECENT STUDIES AND SUMMARY	62
Review of Modern Longitudinal Studies	62
A Modern Study Analyzed by Problem Areas	63
Objectives of the Study	
Selection of Variables and Groups	
Statistics Used	
Attrition	
Cultural Shift	
Results and Conclusions	
Perils of the Problem-oriented Paper	69
Summary	70
A SELECTED BIBLIOGRAPHY	74
APPENDIX	81

LIST OF TABLES

TABLE		PAGE
I	Three Dimensional Table Showing Five Ages of Cohorts when Measured Annually	9

LIST OF FIGURES

FIGURE		PAGE
1	Schematic paradigm demonstrating relationship between individual data gathered over time when evaluated by both analysis of means and of correlation analysis.	11
2	Two logistic curves and their average	12

CHAPTER I

INTRODUCTION

I. PURPOSE OF THESIS

The purpose of this thesis is to review "longitudinal study" in terms of problems and present use. A search of literature on method reveals problems centering around two areas: (1) definition of "longitudinal study" and (2) practical problems of method itself.

The purpose of this thesis is then to explore the following questions:

1. How can "longitudinal study" be defined?
2. What problems are inherent in study of the same individuals over time and how can these problems be solved?

A third question which emerges from these two is:

3. How is "longitudinal study" being used today?

II. HISTORY OF "LONGITUDINAL STUDY"

Kodlin and Thompson (1958) point out that the "idea that certain events and characteristics 'late' in life are dependent on events or characteristics 'early' in life" is deeply embedded in our culture in early writings. In the mythical concept of Moira the Greeks expressed the idea that a person's fate was determined by the spinning of a thread on the day of his birth. In our language we find expressions such as, "the child is father to the man."

Systematic observation of this concept was not attempted until late eighteenth century. The idea may be traced from the "baby biographies" through cross-sectional studies of early physical anthropometrists in late nineteenth century to the first "longitudinal studies" recognizable as such. Dearborn and Rothney (1941) report a study by H. P. Bowditch of Harvard Medical School in 1872. Bowditch measured height of 12 males and 12 females annually for 25 years. In the early part of this century B. T. Baldwin was one of the first to recognize the importance of the "longitudinal" approach. (Kodlin and Thompson 1958)

In mid-twenties the National Research Council held a series of conferences to discuss problems of "the whole child." One outcome of these conferences was that a number of institutions established long-term, multidisciplinary studies. In 1964 Jerome Kagan published a review of 10 of the largest and best known of these "old" studies.

In 1970 Lester W. Sontag, for 40 years director of Fels Research Institute, presented a paper to the American Psychological Association tracing evolution of longitudinal study from the twenties to the present. According to Sontag they had "elected a methodology without delineating many specific problems." Further, "They were committed to making a variety of kinds of observations and measurements for which there were not yet immediate research questions or design." (Sontag 1971)

Today Sontag describes the status of "old" studies which survive as "thoroughly specific problem oriented." "Infinitely more sophisticated instruments of measurement and, of course, data analyses

makes them bear little resemblance to the child development institutes of 40 years ago." In the interim two ideas which had sparked the beginning of the longitudinal studies, whole child concept and the multidisciplinary approach had been largely abandoned.

Perhaps most recently the usefulness of longitudinal method to the developmental researcher has been reaffirmed at the West Virginia Conference on Life Span Developmental Psychology. In the report of the conference (Goulet and Baltes 1970) J. F. Wohlwill concludes as did the scientists who attended the National Research Council conference almost a half-century earlier: "Longitudinal analysis is the method par excellence of the developmental psychologist, and, in many cases, the only one that will permit the requisite detailed analysis of developmental functions."

III. NEED FOR REVIEW OF "LONGITUDINAL STUDY"

A number of summaries and reviews of "longitudinal study" have appeared in the past 25 years. (Tanner 1947; Kodlin and Thompson 1958; Stone and Onqué 1959; Goldfarb 1960; Kagan 1964; and Sontag 1971) Some reviewed subject matter, others method and some attempted both. One of the first of these by Tanner (1947) summarized longitudinal projects, particularly those related to physical growth. In 1958 the Society for Research in Child Development published a monograph by Kodlin and Thompson entitled "An Appraisal of the Longitudinal Approach in the Study of Growth and Development." Here statistical analyses clarify several basic problems of longitudinal method.

In 1950 in the first issue of the Annual Review of Psychology H. E. Jones and N. Bayley said: "Except in a few research centers . . . psychologists are as yet making little systematic use of the longitudinal approach." In the mid-fifties to reawaken interest in longitudinal study and to demonstrate usefulness of the approach Stone and Onqué (1959) reviewed nearly 300 studies.

Although Stone and Onqué report many studies in the late thirties and early forties (23 in 1938), by 1953 number of studies reported had dwindled to 8 and in 1954 only 4 studies are reported. Methodological problems had become so apparent by mid-century that not only were results of these studies questioned but the method itself.

Some studies reviewed in Stone and Onqué do not appear by modern standards to be scientific research. Of the first 30 of these studies 22 apparently had no control groups. Numbers of subjects were often small. Thirteen studies in the total 300 had only one subject in spite of the authors' attempt to eliminate such studies. For some of the longer studies problems of planning, attrition, and cultural shift were especially difficult.

A study started in the twenties was necessarily planned in terms of scientific method as conceived at that time. Inadequacies of method and statistical techniques sometimes did not become evident until data was processed decades later. Apparently, while experimental method generally was being adapted to psychology as a useful tool, longitudinal method remained in confusion and little understood, an anachronism rooted in the 1920's. Even today Sontag (1971) remarks: "The term 'longitudinal research' . . . implies to some scientists a sort of

'womb to tomb' research plan, replete with inadequate design, inexact instruments of measurement, and low-research product."

Although Stone and Onqué did not deal with problems of method, Goldfarb (1960) was concerned with them. His is one of the few works devoted entirely to longitudinal method. Overall point of view leaves much to be desired, however, particularly for the developmental psychologist.

A number of researchers include in their studies résumés of longitudinal studies related to their own work. (Bloom 1964; Robins 1966) Currently, a government bulletin Research Relating to Children lists on-going longitudinal studies and publications resulting from these studies as well as new studies being started.

Several early longitudinal researchers (Dearborn and Rothney 1941; Jones 1958; Sontag 1971) have written short descriptions of longitudinal method. They have described many of their own problems with the method, sometimes with much candor. Whatever the problems of method each new generation of developmentalists rediscover that longitudinal study is necessary to attain answers they need; however paradoxical the problems may seem, they cannot be ignored.

Each of the works mentioned here has made some contribution to solving the dilemma of longitudinal study. As early as 1958 Kodlin and Thompson reported: "Review of the major longitudinal growth studies in this country reveals that increasing attention is being given to improving methodology of these studies." The number of recent studies which investigate special problems of longitudinal method confirm that improvement of longitudinal method is still a live issue today. None

of these works, however, has treated the three purposes of this thesis except in a superficial way.

Chapter II will define longitudinal study and differentiate it from other related methods. Chapter III will treat problems especially relevant to longitudinal study and will report studies which investigate these problems. Chapter IV will review recent longitudinal studies in terms of problems described in Chapter III and will include a short summary of this paper. Particular problem areas which need further research or new problem areas will be noted here.

CHAPTER II

"LONGITUDINAL STUDY" DEFINED

One source of confusion with regard to longitudinal study is in defining the word "longitudinal." One dictionary definition of "longitudinal" in reference to scientific study is simply a "study over time." According to this definition any study where time is involved would be a longitudinal study. In keeping with this definition every conceivable method of study has been called a "longitudinal study." Cross-sectional studies, experimental studies, observational studies, retrospective studies, and studies from records in the past, all have from time to time been referred to as "longitudinal" when time was variously a factor in their design. Failure to differentiate clearly between these many methods of study, their special problems and best uses has been a major factor in confounding the problem of defining "longitudinal study."

Most researchers (Stone and Onqué 1959; Kuhlen 1951; Kodlin and Thompson 1958) who have attempted to define longitudinal study restrict it not only to a study over time but also to a study of the same individuals measured repeatedly. According to Kodlin and Thompson: "The term longitudinal studies is used . . . to designate those studies in which repeated observations over a period of time are made on the same individuals."

I. RELATIONSHIP OF "LONGITUDINAL STUDY" TO CROSS-SECTIONAL STUDY

After defining "longitudinal study" in this way Kodlin and Thompson (1958) add: "We exclude from this term those studies in which repeated observations over a period of time are made on groups or samples of individuals without concern as to the inclusion of the same individuals in each series of observations." They are, of course, referring to cross-sectional study over time. Of cross-sectional and longitudinal studies Baltes (1968) remarks: "Despite the importance of these conventional designs for developmental psychology, it is surprising that nowhere in the psychological literature a satisfactory comprehensive discussion can be found."

Cross-sectional, Time-lag, and Longitudinal Studies

In a 1965 Psychological Bulletin article K. Warner Schaie demonstrated a general statistical model for study of developmental problems. Schaie showed that three types of study, conventional cross-sectional, cross-sectional over time, and longitudinal, are special cases of a more general model. Schaie pointed out elements of these studies are often ignored:

. . . Kessen (1960), in his excellent discussion of research designs for developmental problems, suggested that the investigation of simple age functions is naive and proposed that most meaningful designs involve either the interaction of age and environmental changes or the analysis of age functions in special populations. Only in passing did he mention that even more typical case where aging interacts simultaneously with differences in population and environment.

Table I shows components of Age, Population (Year of Birth or Cohort),

TABLE I*
THREE DIMENSIONAL TABLE SHOWING FIVE AGES
OF COHORTS WHEN MEASURED ANNUALLY

Year of Birth (Cohort)	Age									
1960	6	7	8	9	10					
1961		6	7	8	9	10				
1962			6	7	8	9	10			
1963				6	7	8	9	10		
1964					6	7	8	9	10	
Time of Testing	1966	1967	1968	1969	1970	1971	1972	1973	1974	

*This table modified and simplified from Schaie (1965). The age groups marked horizontally represent a possible longitudinal study; groups marked vertically represent a possible cross-sectional study; and groups marked diagonally represent a possible time-lag study.

and Environment (Time of Testing) in conventional cross-sectional study, cross-sectional study over time and longitudinal study. Schaie referred to the cross-sectional study over time as the time-lag study, and for convenience it will be so designated in this study to avoid confusion with conventional cross-sectional study.

Age represents those characteristics attributable to developmental stage; Year of Birth or Cohort represents the special population or history of events which each individual shares with other individuals born at the same time; and time of Testing represents environment and

events which may be acting upon all persons measured at time of testing. Any one of these three elements may be ignored in longitudinal or cross-sectional studies only when it can be demonstrated that element is not a significant variable for the particular question at hand.

We can easily see that these elements are far from independent. Age is much involved with both Year of Birth and with Time of Testing. Year of Birth and Time of Testing are both aspects of the phenomenon we call "cultural shift."

Longitudinal study measures age changes; Year of Birth (Cohort) is held constant while Age and Time of Testing vary and are confounded. The risk here, therefore, is in attempting to generalize age differences over time from one cohort to another in this type of study. Cross-sectional study measures age differences; Time of Testing is held constant while Age and Year of Birth (Cohort) are confounded. Wohlwill discusses problem of cohort confounding in cross-sectional study at some length. In this type of study the risk is in generalizing Age differences from one Time of Testing to another. Time-lag study measures cultural change; Age is held constant while Year of Birth and Time of Testing vary and are confounded. Needless to say, therefore, we cannot generalize from one age to another in time-lag study.

Cross-sectional, time-lag, and longitudinal studies cannot produce comparable results since they by definition can never measure the same population with its varying components of Age, Year of Birth, and Time of Testing. Puzzling discrepancies between results of cross-sectional and longitudinal studies may be partially explained in this way.

Individuality of Data Retained in Longitudinal Study

Individuality of data derived from testing the same individuals repeatedly is a feature of the longitudinal study which cross-sectional studies do not have. Unless this individuality of data is retained much of the effectiveness of longitudinal study is lost. Correlational analysis is used to retain individuality of data in most longitudinal studies. (Reese and Lipsitt, 1970)

Correlational procedure measures degree of covariation between two variables. This statistical procedure is represented visually in Figure 1. Difference between each individual's score at one testing and a later testing is an integral part of correlational analysis. Such direct comparison of scores is not possible when an average is computed for a whole group and compared with an average for another group. Although means may be compared, individual variations in the relationship

	Experimental Group	Control or Comparison Group
Analysis when means are used:		
Measurement 1	μ_1 x x x x x	μ_2 x x x x x
Measurement 2	μ_3 x x x x x	μ_4 x x x x x
Correlational analysis:		
Measurement 1	r $\left\{ \begin{array}{l} x x x x x \\ x x x x x \end{array} \right.$	r $\left\{ \begin{array}{l} x x x x x \\ x x x x x \end{array} \right.$
Measurement 2		

Figure 1. Schematic paradigm demonstrating relationship between individual data gathered over time when evaluated by both analysis of means and of correlation analysis.

between early scores and later scores are not considered. A critique of correlational analysis is found in Chapter III of this thesis.

Analysis using means may be used to good effect to supplement correlational analysis in longitudinal study adding another dimension of information but not replacing individual information contained in the correlation. Wohlwill (1970) points out two instances where use of means distort individuality of longitudinal data. "Rate of change and above all, shape of the developmental function are examples of such temporal-pattern data subject to distortion through group averaging." Wohlwill presents a "particularly convincing illustration of such distortion" taken from Merrill (1931). Figure 2 shows two individual curves which are averaged to produce a mean curve markedly different from individual curves. Wohlwill notes further:

. . . Note that in this instance the composite averaged curve appears more complex than the component individual curves; a more typical result, however, is that averaged curves result in the smoothing out, not only of random irregularities in the data, but of systematic departures from simple form . . . so as to lead to a spuriously simple form.

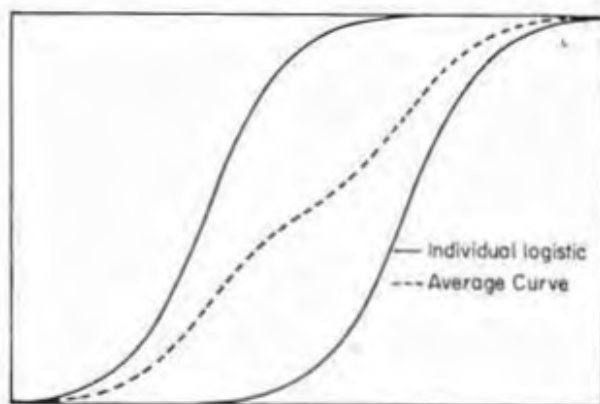


Figure 2. Two logistic curves and their average. (From Merrill 1931)

Bimodal distributions are an example of such systematic departures. Wohlwill modifies this quotation, however: "In the study of behavioral development this particular problem is probably less serious than it has been considered to be either in the study of learning . . . or physical growth . . ."

Schaie (1965) uses means in a novel analysis of variance technique in his mathematical model and finds that data gathered longitudinally adds nothing that cannot be duplicated by the cross-sectional methods. Kodlin and Thompson (1958) also demonstrate in a mathematical analysis that when means are used, much of the individuality of data is lost. "In other words, completely longitudinal records are of no advantage over cross-sectional ones if they are only used to estimate means at particular occasions, and 'mean growth curves' established in either way are equally informative." When individuality of data is not important and averages of groups only are needed, longitudinal method is no longer a matter of necessity but a highly individual choice between longitudinal or cross-sectional methods depending on characteristics of the study at hand. Problems of selecting a method of study are considered further under Planning the Study in Chapter III.

Summary of Cross-sectional Study

Relationship between cross-sectional and longitudinal study may be summarized with this statement from Kodlin and Thompson (1958):

1. The longitudinal approach is the only approach which gives a complete description of the growth phenomenon.
2. The cross-sectional approach never can satisfy the objective of a study which requires the measurement of the change in a trait through time on a given

individual. This means that when the objective of the growth study is to arrive at predictions of individual growth, generally, or, to establish the correlation between measurements obtained at successive ages, it is necessary to employ the longitudinal approach.

Comparison of cross-sectional and longitudinal studies shows that not only is longitudinal study a study over time in which the same individuals are measured repeatedly, but also that individuality of data obtained in this way must be retained. Correlational analysis is the preferred technique when individuality of data must be retained.

II. RELATIONSHIP BETWEEN LONGITUDINAL AND EXPERIMENTAL STUDIES

Thorndike's Study

In the early part of this century experimental studies were not considered feasible for developmental psychology. (Andrews 1948) In 1924, however, at a time when the "big" longitudinal studies were being initiated, E. L. Thorndike published his classic study of transfer of training.

Students were matched according to type of curriculum the student followed, i.e. English, history, geometry and Latin or English, history, geometry and shopwork. Thorndike measured student's intelligence and let the independent variables, School Subjects, operate. After one year he remeasured intelligence. Kerlinger (1964) considered this study a landmark. Although results were far from conclusive, and the design itself was experimental only in a borderline sense, Thorndike's study showed that experimental design could be used in child psychology and pointed the way toward other experimental research in

child psychology.

Obviously, many experimental studies would qualify as longitudinal studies in the usually accepted definition, i.e., a study over time in which the same individuals are measured repeatedly. Stone and Onqué (1959) review a number of experimental studies over time in their monumental "round up" of longitudinal studies.

Longitudinal Study as Study of "Things as They Are Naturally"

Kerlinger (1964) considers "longitudinal study" only as a particular experimental time design. Sontag (1971) considers "longitudinal study" either "experimental or natural." He, thus differentiates the traditional longitudinal studies from experimental studies over time. The "old" studies initiated in the 20's were studies of "things as they are naturally." Experimental study is by definition a contrived or controlled study. The experimenter does not wait for natural occurrence of the relationship he wishes to study but creates such an occurrence to study at his own convenience. Unfortunately many behaviors of interest to the developmentalist cannot be duplicated in an experimental design.

Comparison of Experimental and Longitudinal Design

According to R. A. Fisher all variables which are not completely controlled must be allowed to operate with complete freedom. (Johnson 1949) Needless to say studies which completely fulfill this assumption are rare. Instead we may imagine a "freedom-control continuum" of studies. At the control end of such a continuum we may imagine some of the experiments of Skinner in which virtually every aspect of both subject (rat) and environment (Skinner box) are under control of

the experimenter.

At the freedom end of such a continuum we may find a natural longitudinal study which has been set up in such a way that behavior and environment of the subject are not interfered with in any way by the fact of experimenter and experiment. In this situation all knowledge of the experiment would preferably be kept from the subject. Such experiments are possible in many school or penal settings where a structured or controlled environment is the natural setting for the individual. Most studies both experimental and longitudinal fall somewhere between these extremes. Depending upon the ingenuity of experimenter "controlled" variables do not interfere with "free" variables in both experimental and longitudinal studies.

In the experimental study, experimental and control groups are as much alike as possible. Researcher treats the experimental group in some way differently from the control group. In other words, the experimental variable is manipulated. Subsequent difference between dependent variable of interest in the two groups is assumed to be due to experimental treatment.

As Goldfarb (1960) points out, traditional longitudinal study does not include manipulation of the independent variable in the experimental sense. Since longitudinal study is study of a natural situation, the experimental group is selected on the basis of some important variable to be studied (independent variable). Only in this sense is the independent variable manipulated.

After experimental and control groups are differentiated, the researcher observes individual behavior and takes every precaution

not to control behavior in any way. Expectation in a free and uncontrolled situation is that variables other than the independent variable will statistically cancel each other unless they are in some way related to the independent variable for which groups were originally selected. Thus, longitudinal study using correlation analysis will show co-variation between independent variable and another variable.

Experimental study may also differentiate groups according to some independent variable as in longitudinal study, i.e., in analysis of variance design where groups may be differentiated according to sex, age, grade, etc. For both types of study this feature introduces a problem of randomization which is further discussed in Chapter III.

Interpretation of "Causality" in Experimental and Longitudinal Studies

Campbell and Stanley (1969) discuss the relative merits of some experimental and correlational designs to show "causality." They say: "[Correlational data] are relevant to causal hypotheses inasmuch as they expose them to disconfirmation. If a zero correlation is obtained, the credibility of the hypothesis is lessened." They add:

A perusal of research on teaching would soon convince one that the causal interpretation of correlational data is overdone rather than underdone, that plausible rival hypotheses are often overlooked, and that to establish the temporal antecedence-consequence of a causal relationship, observations extended in time, if not experimental intrusion of X, are essential.

Of experimental results they say:

In a very fundamental sense, experimental results never "confirm" or "prove" a theory-- rather, the successful theory is tested and escapes being disconfirmed. . . .

It is by now generally understood that the "null hypothesis" often employed for convenience in stating the hypothesis of an experiment can never be "accepted" by the data obtained; it

can only be "rejected," or "fail to be rejected."

Due to fallacy of inductive logic "causality" remains in doubt in any scientific experiment.

Campbell and Stanley further caution that all experiments must be interpreted according to the limitations of the particular design used. As an aid in designing as well as interpreting experiments they have constructed several tables showing plus and minus values of a number of experimental designs. Several of these are correlational.

Predictive Usefulness of Longitudinal Study

Granted that longitudinal data must be interpreted with care and is not always as effective in "eliminating plausible rival hypotheses" as experimental results, longitudinal studies have often predictive usefulness in an area that the experimentalists cannot penetrate.

Prospectus on Research, White House Conference on Children and Youth (1960) reports many requests for "intensified research in development of predictive devices for early identification of children likely to engage in serious or persistent delinquent behavior." Predictive studies are also needed in the area of school dropout, vocational guidance, and welfare recipients as well as a need to predict creative, productive members of society.

Summary of Relationship between Longitudinal and Experimental Studies

Comparison of longitudinal and experimental studies further defines longitudinal study as a study of things as they are naturally as opposed to the controlled experimental study. Ideally in longitudinal study all variables operate with complete freedom. Longitudinal

and experimental studies differ in selection of groups, manipulation of independent variable, and degree of control of extraneous variables.

III. RETROSPECTIVE AND PROSPECTIVE STUDIES

Kodlin and Thompson (1958) dichotomize longitudinal study as "prospective" and "retrospective." The prospective longitudinal study is the traditional study from present to future which will be defined in this thesis. In retrospective study cross-sectional samples are questioned about events in the past. Kodlin and Thompson find retrospective study mathematically feasible substitute for prospective longitudinal study but greatly hampered by fallacy of memory. Not only do people simply fail to remember, but memories become distorted over time; when outcomes are known, subjects have a tendency to rationalize antecedents. (Kessen 1960; Thomas et al 1963) Subjects are all too willing to supply experimenter with a plausible ex post facto "explanation" of some present behavior unless hypotheses are well researched in advance and concisely presented. (Campbell and Stanley 1969)

Nonetheless, retrospective study has several advantages. Data collected in the present does avoid to some extent many problems of traditional longitudinal study, i.e., cultural shift and attrition. Like cross-sectional study to which it is also related retrospective study can be used when time or facilities for longitudinal study are not available. The researcher has the added advantage of being able to select indices which are known to be appropriate in the present.

Yarrow, Campbell, and Burton Study

Yarrow, Campbell, and Burton (1970) compared data from three sources: records, retrospective data, and data from a short longitudinal study. The retrospective study questioned mothers and children traced from the records. The authors intended the study "to examine consistency between different sets of data on the same subjects--gathered at different points in time, by different methods, and from different types of informants."

They summarized a number of earlier methodological studies of retrospective research. Attitudinal items and areas associated with subjects anxieties showed least agreement. Retrospective presentations tended to bias report in desired direction; sex and ordinal position of child seemed to affect data; mothers seemed to be more accurate in some areas of recall than others. Their own study showed ability to remember was affected by type of question, wording of question, and other techniques of data gathering.

In a subsequent article Burton (1970) demonstrated multitrait-multimethod matrix analysis using the same data. Results were the same, but the presentation more orderly than study using more conventional methods of analysis. The matrix showed that all possibilities of data had been explored.

In the Yarrow, Campbell, and Burton study records were not of good quality, and attrition seemed unnecessarily large. This study pointed the way, however, toward some possibilities for improving retrospective method. The advantages of simply asking subjects for information are so great that further methodological research to

improve quality of retrospective data is much needed.

Summary of Prospective and Retrospective Study

The present analysis of retrospective method shows it is not "just a substitute" for longitudinal study. Retrospective method and the study from records to be discussed in the following section investigate a dimension not available to any other types of study, the past. Advantages and problems of retrospective method are seen to be very different from longitudinal method. By comparing prospective and retrospective studies longitudinal study is further defined now as the study which looks to the future.

IV. STUDY FROM PAST RECORDS

Retrospective method and study from past records are our only means of exploring the past. Although study from records is sometimes confused with both retrospective study and longitudinal study, it is a distinct method with its own techniques, advantages, and problems.

Advantages of Study from Records

In the study of retrospective method by Yarrow, Campbell, and Burton (1970), cited in the previous section, subjects traced from past records were asked to remember events in retrospect. In the study from records, subjects are usually asked about events in the present. Thus, one of the chief advantages of study from records is that problems of memory are minimized.

Most records are not of good quality and not suitable for scientific purposes. When good records are located, this type study

has several further advantages. By using a "lucky cache" of unusually complete files Robins' (1966) follow-up study of delinquents reduced to a minimum problems of conventional longitudinal research, i.e., cultural shift, changing personnel, and long wait for results. Although collection and analysis of data did take several years and a number of coworkers, continuity of design, philosophy, and set of interests" was maintained throughout.

Problems of Study from Records

Ex Post Facto Research. In any study order of scientific procedures are important, but in study from past records particular care is taken that hypotheses are made before data are analyzed, and records from the past are examined without reference to data from the present. To insure that the study not become an ex post facto research in Robins' study null hypotheses were made, data from the present collected without reference to previous data, statistical analyses made and conclusions drawn in usual order of a scientific study.

A number of techniques are used to prevent knowledge of outcome of subjects from influencing hypotheses and design of study. Files may be coded to avoid recognition of individuals; researcher may conceal identity of controls and experimental subjects; someone not familiar with work in the present may be brought into the study to analyze data from the past.

Hypotheses: Theoretical or Fortuitous? Records are relics from the past. They cannot be changed; they cannot yield additional information; researcher is circumscribed by what is there. Under these

circumstances the question arises: To what extent are hypotheses dictated by availability of records rather than theoretical orientation? On this point Robins (1966) said:

. . . Research is supposed to follow a classic model: First, develop a hypothesis; second, find a way to test it. This study is an example of the fact that research in practice seldom arises from a full-blown hypothesis in search of a crucial test. . . .

. . . The means in this case largely determined our research ends, although there were many other problems which the same population would have allowed us to study.

Thus, many questions are latent in most areas of psychology awaiting only an opportunity to examine them. In like manner preponderance of studies of college students is not dictated altogether by a desire to know more and more about the college population. Availability remains a determining factor in research whether records or subjects are used.

Attrition. When records are traced to individuals in the present, attrition may be a major problem. For Robins the problem of locating subjects after 30 years proved an expensive one, but with a well-developed and feasible research plan to present, funds were made available from government grants and private research funds. The low attrition rates Robins achieved through having funds for travel and interviewers are discussed further under Attrition in Chapter III.

Ethical Problem of Use of Records. An important problem of study from records and one least amenable to solution at this time is the ethical question of whether researcher has the "right" to use records. Several sources (Eisenberg in Robins, 1966; Bloom 1964) enthusiastically support use of records as a source of data. Bloom

advocates that records from schools and other social agencies be planned with research as a goal. D. T. Campbell (1969) discusses this problem of use of records within the framework of the larger problem of any psychological experiment which injures an individual through use of "potentially embarrassing material" or "humiliation and insult."

. . . On the other hand, in those settings using means and ranges of communication that are within the public domain, and which non-experimenters are using freely without such permission, this [request for permission] becomes an utterly unreasonable requirement. (Campbell 1969)

Subjects permission to use records may be sought, but if it is not granted, the problem of attrition is compounded.

The ethical question involved in use of public records seems to revolve around several questions. How public are public records? How public should public records be? These two questions are further confounded with the question: Is it possible to keep public records private?

At the present time our more defenseless citizens are subject to invasion of privacy depending on judgement and consideration of keepers of the public archives. These citizens include school children, criminals, delinquents or those suspected of crime or delinquency, the mentally ill or retarded, welfare recipients, the elderly, physically handicapped, members of the armed forces, etc. Some citizens not usually considered helpless may find records they consider private, made public, i.e., college students whose records may be examined, persons involved in litigation, i.e., divorce, as well as those who pay property or income taxes.

Reese and Lipsitt (1970) review ethical considerations with

special reference to research with children. They quote extensively from a statement (Newsletter 1968) on ethical standards for developmental psychologists of the American Psychological Association. In general the researcher is held responsible for the child's welfare. It is further recommended that consent of parent and child be obtained. With regard to records specifically the child's identity is concealed, and the researcher does not report test scores or other information given in confidence to parents.

Summary of Study from Records

Study from records adds no new dimension to definition of longitudinal study. It is simply another way of exploring the past. Its advantages when good data are at hand offer unique opportunities for researchers. Concerted effort to solve its problems would be a worthwhile contribution to research.

V. TRADITIONAL LONGITUDINAL STUDY

"Longitudinal Study" Defined

By comparing longitudinal study with cross-sectional study, time-lag study, experimental study, retrospective study, and study from records, traditional longitudinal study has been further defined. When the same individuals are tested or observed on two or more occasions, the study is necessarily a study over time. Only through change of some kind does time exist, therefore, the study over time is a study of change.

In comparison with cross-sectional and time-lag studies

longitudinal study is seen not only as a study of the same individuals measured repeatedly over time, but also that this individuality is retained in analysis of data. When individuality of data is to be retained, correlation analysis is most commonly used. By retaining individuality of data experimenter may study individual change itself, characteristics of the individuals who change, and rate of individual change. (Kuhlen 1951; Goldfarb 1960)

As compared with the controlled experimental study longitudinal study is defined as the study of "things as they are naturally." Finally, in comparison with retrospective study and study from records longitudinal study is differentiated as prospective study, that is, the study which proceeds from present to future rather than to the past.

A scientific method cannot be defined in terms of one or two facets of the method. A scientific method is a method of sampling, a method of data collection and analysis, method of testing or observing and of procedures generally. Defined more concisely now traditional longitudinal study is the study: (1) of individual change under natural conditions not controlled in any way by the experimenter, (2) which proceeds over time from the present to the future by measuring the same individuals repeatedly, and (3) which retains individuality of data in analyses.

"What's in a Name?"

All types of studies over time have been referred to as longitudinal studies. Some researchers consider longitudinal study only as the experimental study. (Kerlinger 1963) Others consider longitudinal

study an equally appropriate title for either the experimental study or the traditional study. (Sontag 1971) Comparison of natural and experimental studies over time showed that while the two types of studies are much alike in many respects, design and problems are also basically different. Traditional longitudinal study has been called the natural or naturalistic study, the observational study, prospective study, follow-up study, data analysis, correlational or differential study, non-experimental time series, continuous record, repeated measurement study, etc. All these names are sometimes used for other methods of study. "Natural" or "observational" study is sometimes used to designate simple observations and classifications. (Butler, Rice, and Wagstaff 1963; Willems and Raush 1969) Data analysis is so general a term as to be practically meaningless. In this thesis longitudinal study will continue to be used to designate the traditional study over time as it has been defined here from lack of a more appropriate name.

A Traditional Longitudinal Study

One of the first and most meaningful of the "old studies" was Study of the Gifted Child initiated by Lewis Terman at Stanford University in 1920. Terman (1926) attributes much of the early success of that study to the man who served as statistician for the project T. L. Kelley. In this design the experimental group of 1000 school children was selected on the basis of the independent variable, intelligence. The children were measured on a variety of other variables of interest, both physical and mental. No attempts were made to manipulate or control the children in any way. They were located and

CHAPTER III

PROBLEMS OF LONGITUDINAL STUDY

This chapter will present some problems which are either unique to longitudinal studies or which have proved particularly troublesome. References and recent research related to these problems will be discussed here.

Planning the longitudinal study itself and the related problem of cost are problems of primary importance. The ubiquitous nature of cultural shift and attrition have already been mentioned in Chapter II but will be more fully treated here. Finally, sample selection, randomization, and generalization of the study results present a dilemma for any design where one group is chosen on the basis of an independent variable. This problem is particularly relevant to longitudinal study.

I. PLANNING THE STUDY

In planning a study over time basic scientific methodology and best use of longitudinal method are prime considerations. If the researcher is not an expert in scientific methodology, his first consideration in planning the study is to find such a person available for consultation.

Definition of Study Objective.

Hypotheses and objectives of study are clearly defined regardless of method of study according to Kodlin and Thompson (1958). They emphasize especially that traits under study be meaningful and of proven validity. Specific objectives and hypotheses are often an outgrowth of preliminary studies. A selection and pretesting of variables may forestall an unproductive study. (Goldfarb 1960)

Exploratory studies may be cross-sectional, retrospective, naturalistic observations and classification, studies of one individual, or even experimental studies or relatively short longitudinal studies. Such studies may be helpful not only in defining the objective of a study but also in anticipating other problems of planning and procedures, i.e., sample selection, cost, etc. (Kodlin and Thompson 1958; Goldfarb 1960)

On the other hand when objectives of the study can only be satisfied by longitudinal method, preliminary studies may be misleading. The researcher may become discouraged with negative results of studies which cannot give satisfactory answers to questions of interest.

Selection of Method of Study.

After objectives are defined, the method of study is selected. Developmental problems are no longer seen as amenable only to longitudinal investigation. According to Sontag (1971): "The modern human development institute is a challenging and stimulating organization, using a variety of scientific approaches, human and animal research material, with a talented staff fully aware of the advantages they

enjoy." Kodlin and Thompson (1958) review considerations in selection of method of study in detail.

Methods of study have been differentiated in Chapter II. These differences will be summarized briefly here with several additional considerations especially relevant in planning a study.

Cross-sectional Study. If averages or norms are enough to satisfy the inquiry, longitudinal and cross-sectional data will be equally informative under assumption of a stable universe. For the same precision, however, cross-sectional samples must be ten times as large as longitudinal samples. (Kodlin and Thompson 1958) In spite of this disadvantage cross-sectional method may show results sooner.

Experimental Method. If experimental study can measure the variables of interest in a controlled setting, complete randomization in sample selection make results more generalizable. Closer approximation to a "causal" relationship may be established. When experimental design over time is substituted for natural longitudinal study, however, the subject of study may shift subtly. Results may show only what will happen under particular experimental conditions rather than under natural conditions which the researcher wishes to investigate.

Combinations of Methods. In selecting the study method the researcher may consider relating several studies using different methods in such a way that various types of results can be brought to bear on the problem of eliminating alternative hypotheses. (Goldfarb 1960; Mainland 1952) The study of mothers and young children by Yarrow, Campbell, and Burton cited in Chapter II of this paper is an example of an effort to combine information from a study from

records, a retrospective study and a longitudinal study.

Page and Pinango (1970) demonstrate a way in which information from a number of studies may be combined in a matrix. They show on a theoretical level that as number of studies is increased, information is increased exponentially. Rees and Palmer (1970) combine comparable data on factors related to change in mental test performance from five longitudinal studies from the Fels Research Institute, the Child Research Council at Denver and three studies at University of California: the Berkeley Growth Study, the Berkeley Guidance Study, and the Oakland Growth Study.

Many writers have mentioned the desirability of a design which combines the cross-sectional and longitudinal methods, but few concrete proposals have been made for accomplishing such a merger. In 1953 R. Q. Bell proposed a design "combining the cross-sectional and longitudinal techniques in such a way that developmental changes for a long period may be estimated in a much shorter period." He explained his method with an example:

. . . a study of changes in a dependent variable such as synonym usage in children's word definitions over a four-year period may be accomplished by measuring each of two age groups three times over a two-year period. Age groups such as eight- and ten-year-olds would be selected so that the final measurement of the eight-year-olds is made at the same age as the initial measurement of the ten-year-olds.

For convenience he called this method a "convergence" design. The two age groups were matched on a number of variables to insure continuity of a curve such as might have been obtained if one group had been followed over a much longer period of time.

Wohlwill (1970a) believes Schaie's design which combines data for all cohorts at all times of testing corresponds formally with Bell's convergence method. Bell's proposal, like Schaie's, defeats one unique advantage of longitudinal method, individuality of data, for he concludes: "If the cumulative effect of an independent variable on a single individual is to be studied, however, the convergence and cross-sectional approaches cannot be used." Also as Wohlwill (1970a) points out, "with most behavioral measures the task of matching individuals to obtain synthetic longitudinal curves is apt to prove extremely difficult."

Correlation analysis.

As demonstrated in Figure 1 in Chapter II of this paper, individuality of the data obtained in a longitudinal study is retained by correlation analysis. As demonstrated in Figure 2 analysis using means does not always show individual differences accurately. Cronbach (1957) differentiates experimental and correlation analysis in this way:

Individual differences have been an annoyance rather than a challenge to the experimenter. His goal is to control behavior, and variation within treatments is proof that he has not succeeded. Individual variation is cast into that outer darkness known as "error variance."

On the other hand, he adds:

The correlational psychologist is in love with just those variables the experimenter left home to forget. He regards individual and group variation as important effects of biological and social causes.

Wohlwill (1970a) also discusses relative merits of experimental and correlation analysis.

As pointed out earlier, modern philosophy of science shows the impossibility of "proving causality" by inductive, scientific method. We "accept" a hypothesis when we can make a strong case for it. Newton's theories were accepted for many years in spite of known variations in calculations. Einstein's theories have supplanted those of Newton on the basis of more accurate calculations. These calculations based on Einstein's theories are nonetheless still subject to some variation. The possibility still exists that a more accurate theory will be proposed. (Campbell and Stanley 1969)

As Cronbach (1957) points out: "The well-known virtue of the experimental method is that it brings situational variables under tight control. It thus permits rigorous tests of hypotheses and confident statements about causation." These are, of course, relatively confident statements. Correlational statements are relatively weak in the direction of "causality." According to G. G. Thompson (1962):

. . . Correlational analysis is beset with one serious inferential hazard . . . , [for] . . . two variables may be highly correlated with each other on the basis of a covarying (but unknown and unmeasured) third variable. For example, the high correlation between a group-administered intelligence test and a test of academic achievement may be due to a common variable of reading ability.

Campbell and Stanley (1969) describe a further complication:

. . . Even when in a natural setting, an inherent temporal priority seems to be involved, selective retention processes can create a causality in the reverse direction. Consider, for example, possible findings that the superintendents with the better schools were better educated and that schools with frequent changes in superintendents had low morale. Almost inevitably we draw the implication that the educational level of superintendents and stable leadership cause better schools. The causal chain could be quite the reverse: better schools (for whatever reasons better) might cause well-educated men

to stay on, while poorer schools might lead the better-educated men to be tempted away into other jobs.

Campbell and Stanley refer to such alternative interpretations of results as "plausible alternative hypotheses." They point out that elimination of such alternatives strengthens whatever statement the correlation makes relevant to causality. In planning the longitudinal study these alternative hypotheses may be anticipated and relevant data toward eliminating them included in the design. The more "plausible alternative hypotheses" are eliminated, the stronger the statements that can be made from results.

The Lazarsfeld Sixteenfold Table. Campbell and Stanley show another way of strengthening the correlational statement. When data can be dichotomized, the Lazarsfeld sixteenfold table is useful in resolving quandaries such as described by Campbell and Stanley above. They describe this table in some detail:

Suppose that on a given occasion we can classify the behavior of 100 teachers as "warm" or "cold," and the behavior of their students as "responsive" or "unresponsive." Doing this, we discover a positive correlation: warm teachers have responsive classes. The question can now be asked, Does teacher warmth cause class responsiveness, or does a responsive class bring out warmth in teachers? While our cultural expectations prejudice us for the first interpretation, a very plausible case can be made for the second. (And, undoubtedly, reciprocal causation is involved.) A panel [longitudinal] study would add relevant data by restudying the same variables upon a second occasion, with the same teachers and classes involved. (Two levels of measurement for two variables generate four response types for each occasion, or 4×4 possible response patterns for the two occasions, generating the sixteenfold table.)

In this example, analysis showed that the teachers had not changed in warmth. Only pupil responsiveness had changed. Under these circumstances relatively strong "causal" statements might be made. If,

however, both teachers and students had changed, the tables would still have shown relative strengths or degree of the various attitudes. In the planning phase of the longitudinal study the researcher may elect to use data which can be dichotomized in order to have the advantage of this method.

Interpretation of the Correlation. At the very least according to Campbell and Stanley correlational data are relevant to causal hypotheses in that they expose them to disconfirmation. They say:

If a zero correlation is obtained, the credibility of the hypothesis is lessened. If a high correlation occurs, the credibility of the hypothesis is strengthened in that it has survived a chance of disconfirmation.

Correlational analysis may be validly interpreted as showing covariation of two variables, and results may be considered predictive. When a variable is significantly related to another variable, the researcher can predict that they will continue to appear either together or successively. The need for such predictions was discussed in Chapter II. Any other interpretation of a correlation would rest with whatever supporting evidence the researcher could muster along the lines we have been discussing.

Factor analysis. Cronbach (1957) declared that the discovery of "partial correlation is the starting place for all of factor analysis," and compares it in importance to exploration of the conditioned-response. In discussing factor analysis Wohlwill (1970a) expresses the need of a broader statistical repertoire felt by many developmentalists:

Measures of absolute change between two occasions, furthermore, provide only a coarse indication of the course of development. For what we are ultimately interested in is the

however, both teachers and students had changed, the tables would still have shown relative strengths or degree of the various attitudes. In the planning phase of the longitudinal study the researcher may elect to use data which can be dichotomized in order to have the advantage of this method.

Interpretation of the Correlation. At the very least according to Campbell and Stanley correlational data are relevant to causal hypotheses in that they expose them to disconfirmation. They say:

If a zero correlation is obtained, the credibility of the hypothesis is lessened. If a high correlation occurs, the credibility of the hypothesis is strengthened in that it has survived a chance of disconfirmation.

Correlational analysis may be validly interpreted as showing covariation of two variables, and results may be considered predictive. When a variable is significantly related to another variable, the researcher can predict that they will continue to appear either together or successively. The need for such predictions was discussed in Chapter II. Any other interpretation of a correlation would rest with whatever supporting evidence the researcher could muster along the lines we have been discussing.

Factor analysis. Cronbach (1957) declared that the discovery of "partial correlation is the starting place for all of factor analysis," and compares it in importance to exploration of the conditioned-response. In discussing factor analysis Wohlwill (1970a) expresses the need of a broader statistical repertoire felt by many developmentalists:

Measures of absolute change between two occasions, furthermore, provide only a coarse indication of the course of development. For what we are ultimately interested in is the

interrelationship among the total developmental functions for a set of variables.

He adds: "To date, virtually the only attempts to deal with this type of question have taken the form of the factor analysis of measures over a series of occasions . . ."

Nesselroade (1970) expands this topic further:

Without involving ourselves in fruitless debate over such issues as the status of derived variables, such as factors, as psychological constructs, let us admit that there are advantages to being able to examine and, possibly, even to manipulate not single variables but complexes or patterned relationships among many variables in concert. This necessarily involves some degree of abstraction, although one may choose not to move too far from hard data.

Nesselroade finds two advantages in working within the factor analysis conceptual framework. "The first is that of finding out what measures 'belong' to the concept; the second is that of determining what measures do not 'belong.'" He concludes:

Techniques such as factor analysis enable one to systematically explore the interrelationships of a number of variables, all potentially of interest to the investigator, in such a way as to detect how subsets of such variables may behave as a coherent unity over different situations or conditions.

Whatever the uncertainties of multivariate approaches at this time they are being vigorously explored as a means of coping with the multidimensional longitudinal process with its dual problems of individuality and change. Several recent works (Harris 1963; Goulet and Baltes 1970; Cattell 1970) describe a number of modern factor analytic approaches which are currently being explored.

Summary of Correlation Analysis. In a presidential address to the American Psychological Association in 1957 Cronbach compared "the two disciplines of scientific psychology," experimental psychology

and correlational psychology. He said:

. . . Correlational psychology, though fully as old as experimentation, was slower to mature. It qualifies equally as a discipline, however, because it asks a distinctive type of question and has technical methods of examining whether the question has been properly put and the data properly interpreted. . . .

In a plea for rapprochement of "the two disciplines" he said: "A true federation of the disciplines is required. Kept independent, they can give only wrong answers or no answers at all regarding certain important problems."

As Campbell and Stanley (1969) emphasize, no mathematical model is perfect. A model can be safely used only when assumptions and risks of the method are recognized and stated. As mentioned previously, data can be interpreted only within the limitations of the design and statistical techniques of a study whether experimental or correlation analysis.

Cost.

Cost is often cited as an important determinant of method of study. (Kagan and Moss 1962; Kodlin and Thompson 1958; Goldfarb 1960) In the past the rule seems to have been that other less expensive and time consuming methods of study be exhausted before longitudinal study is attempted. The struggle of the early longitudinal researchers for funds are well-known. In the affluent society financing is not as difficult to obtain as it was in the 30's when the "big projects" were often plagued with a shortage of funds.

Several modern researchers indicate that cost is no longer a determining factor in choosing a design. Sontag (1971) finds

longitudinal study may be "no more expensive" than cross-sectional design. He affirms "today longitudinal means a very different kind of research than was projected at the time the original longitudinal institutes were conceived." In limiting objectives of project and introducing more precise measurements or observations costs are automatically reduced.

Modern computers have reduced costs of labor for correlational analyses. According to Churchhill (1966):

For most types of growth study data, the cost of computer time for rather complete statistical analysis will be no more than a trivial fraction of the cost of acquiring the data, and it will often be substantially less than that of punching the data into cards or typing the study's final report.

More complicated analyses may now be carried out routinely which were rarely, if ever, used before. (Churchhill 1966; Wohlwill 1970a)

Goldfarb (1960) includes a chapter on computer analyses.

Goldfarb also finds some cost advantages in longitudinal study. When sample selection is a substantial cost, longitudinal study may be less expensive to maintain than to repeatedly select samples for cross-sectional studies. Selection of only one sample reduces sample error as well.

Under certain conditions even today cost may become a liability of longitudinal research. When the study is prolonged and attrition is a major problem, costs of tracing subjects may be high, especially if fund for travel are required to trace subjects. Advance planning may reduce this financial hazard.

Time Factor in Longitudinal Study.

Long wait for results is often considered an overriding objection to longitudinal study. If results are needed at once, this may well be the case. In modern longitudinal study, however, time objectives as well as problem objectives are limited. Length of study is simply long enough for specified change to take place. (Sontag 1970) Even when the ultimate goal is to explore a long period of time, smaller projects with more limited goals may be planned within the larger study. Other researchers (Goldfarb 1960; Kodlin and Thompson 1958) point out advantages of planned reevaluations to keep the project up-to-date in both objectives and procedures.

The time factor is of course vitally involved in the problems of cultural shift and attrition to be discussed subsequently. It is also of practical significance in planning the staff of the study. Psychological significance of time permeates the whole study, i.e., the tendency to accept accidental findings without replication discussed in the next section. These considerations are not usually sufficient to countermand use of longitudinal study. Temperament of the researcher may be the determining factor here.

Post Hoc Analysis and Replication of the Study.

The virtual impossibility of replication of longitudinal study in the usual sense places an added burden on planning the study. Kessen (1960) emphasizes careful selection of variables and relationships to be studied. He notes that in the planning process the tendency to wring all possible information, especially from large, expensive projects

has in the past lead researchers to continue adding one more variable until the project arrived at unmanageable proportions. When a great many relationships are explored in a large project, some results will almost certainly be due to chance even when they are predicted in advance.

The probability that a relationship really exists is statistically greatly enhanced when it is specified in advance. When a null hypothesis is rejected, the researcher may be much more confident statistically that the relationship really exists, than when an apparent relationship turns up a surprise to all. In longitudinal research this principle has not been sufficiently emphasized. An apparent relationship that is discovered by accident may be subject for future investigation but is of little immediate significance.

According to Kessen (1960): "The major safeguard to be built against this danger of 'discovering' merely accidental relationships in post hoc longitudinal analysis is replication or cross-validation." Due to length of study, cultural shift, etc. exact replication of the study is impossible. In the planning stage if enough subjects are available, however, they may be divided randomly into two or more groups and several experiments run simultaneously. Thus, by appropriate planning replication is built into the design of the study.

Problem of Allowing Variables to Operate Freely.

Results may be more clear cut if the difficulty of keeping the study itself from interfering with the free operation of all variables is anticipated in planning the study. Glueck and Glueck (1968) comment

on this problem. Their study showed that delinquents felt that parents did not care for them more often than did controls. In a study of delinquents the researcher continues to turn up from time to time and seems to care a great deal about the subject. How can this influence in the subject's life be assessed or avoided? The Gluecks point out that even controls may have difficulty explaining to wife and children the presence of an interviewer who suddenly comes asking questions.

As with experimental study where the ideal of complete control is seldom possible, this ideal of complete freedom of variables may not be possible. A number of studies show that reasonable anticipation of this problem in the planning stage may minimize this artifact. Several aspects of this problem are discussed more fully in Artifact in Behavior Research. (Rosenthal and Rosnow 1969)

Summary of Planning the Study.

Planning the longitudinal study involves observation of all the usual scientific protocols. The objectives of the study are defined and hypotheses specifically stated. Sampling, measurements, and other procedures of the study as well as statistical methods are anticipated as for any scientific study. In addition to these considerations in planning the longitudinal study the experimenter weighs cost and time factors and if possible builds replication of the study into the design. He anticipates cultural shift, attrition, and other problems of longitudinal study described subsequently in this chapter, and throughout the planning process he keeps constantly in mind that the study must not encroach upon the natural environment of the subject.

The remainder of this chapter is devoted to discussion of problems especially pertinent to the longitudinal study, cultural shift, attrition sampling, randomization, and generalizability of the study.

II. CULTURAL SHIFT

Age-Time of Testing Confound.

In Chapter II "cultural shift" is seen as inherent in both longitudinal and cross-sectional studies due to confounding of the three components, Age, Time of Testing, and Time of Birth (Cohort). Schaie (1965 1970) makes it very clear that the broader category which we refer to as cultural shift stems from two factors: Time of Testing and Time of Birth (Cohort). In longitudinal study Time of Testing is confounded with Age. Differences in Time of Testing represent "differences due to commonly experienced environmental variables or changes in cultural context." (Schaie 1965)

Behaviors which are most affected by Time of Testing effects ordinarily remain stable once they are acquired but may be modified due to environmental changes. Word usages, attitude patterns, occupational choices, modes of dress, etc., are examples. (Schaie 1965) Time of Testing effects may be due to a wide range of events, i.e., simply to difference between morning and afternoon measurement or an event such as declaration of war on the day of testing. According to Schaie (1965) most acquired behaviors with maturational sequences unrelated to genetic changes are affected when response is a function of confounded Age and Time of Testing.

Robins (1963) in the study from records, cited previously, reported subtle shifts in the use of words from the time records were made and thirty years later when subjects were tested again. Baltes and Reinert (1969) report a longitudinal study of intelligence of children in the public schools in Saarbruecken. Time of Testing effects were apparent both for season of the year and time of the school year. These effects had been attributed to interaction between age and cohort, but further analysis revealed their Time of Testing origin. Differences due to Time of Birth are a problem for the longitudinal study, not as affecting results directly, but as they affect generalizability of results. When only one cohort is tested, results cannot be generalized except to other members of that cohort unless it can be shown that cohort differences are not involved.

Age-Time of Birth Confound

Differences between Cohorts (Time of Birth) represent "the effect of different life experiences previous to the time of measurement or genetic changes from one cohort to the next." (Schaie 1965) Results which show response to be entirely a function of cohort differences "would mean that there are no cultural or age changes but that response differences must be attributed solely to variation among generations. This situation will arise rarely, except where true genetic changes in physical attributes may be observed." (Schaie 1965)

When response is a function of confounded Age and Cohort (Time of Birth), according to Schaie "most motor skills and perceptual variables which have maturational sequences and show some genetic shift, but

which are unrelated to cultural change," are most affected. Kessen (1960) cites a hypothetical example of this condition in a cross-sectional study:

. . . a cross-sectional study of night fears or response to sudden awakening among London children might well show striking changes at 15 to 18, not necessarily because of some characteristic of late adolescence, but because these children had been through the Battle of Britain at susceptible ages.

A longitudinal study in which Cohort (Time of Birth) is held constant would not have shown such a difference. This is not to say that such an effect may not be influencing the particular cohort studied longitudinally. As pointed out in the previous section such effects limit generalizability of the study. In like manner in cross-sectional study unless Time of Testing can be shown not to have affected results, generalizability of the study is limited to that Time of Testing only. Generalizability of the study will be discussed more fully later in this chapter.

Schaie describes results of longitudinal and cross-sectional studies:

. . . It should be noted that the confounded Ad + Cd [Age differences + Cohort differences] and Ad + Cd [Age differences + Time of Testing differences] terms accurately represent what are conventionally described as cross-sectional and longitudinal age changes.

He has constructed an interesting table showing comparison of inferences drawn from cross-sectional, longitudinal, and short-term longitudinal studies. The latter uses Schaie's design in which confounding of cultural shift with age gradients is minimized. Variables measured were verbal meaning, space, reasoning, number, and word fluency. Cross-sectional and longitudinal results varied widely sometimes

showing gains and sometimes decrement without apparent consistent pattern. The short-term longitudinal study showed a consistent pattern of modest decrement of the first four variables varying in onset from young adult plateau to advanced age. Word fluency showed sharp decrement from young adult levels. Cross-sectional study had showed moderate decrement and longitudinal study showed moderate gains on that variable. Differences in results of these three types of study are presumed to be due to cultural shift confounding.

Schaie's model is appropriate only when analysis of variance results will give the answers needed. It is introduced here only to show the sharp data differences when net age measures are not attained. A design is not yet available which also retains individuality of data.

Strategies.

Welford (1958) describes the problem of measuring cultural shift in the longitudinal study:

What is probably the most recalcitrant and all-pervading methodological problem for aging studies derives from the fact that a human being carries his history with him. Each new situation is dealt with in terms brought from past experience and in its turn modifies the experience brought to bear on future situations. The organism is thus constantly changing; its present state cannot be fully understood without some reference to past events, and we can never assume that it is the same on two different occasions-- indeed we can be sure it is not.

Thus, the problem of longitudinal study is of measuring an on-going process rather than a static situation. Several strategies have been suggested for coping with this amorphous problem.

These strategies range from simple common sense solutions to data analysis techniques so complicated that they were not feasible until the advent of computerized analysis. As an instance of a simple solution, the Time of Testing confounding in the Baltes and Reinert (1969) study of school children might have been avoided by measuring the children at one season or at one time of the school year. This obvious technique of course deals only with those two possible confounds and does not avoid all other possible Time of Testing effects.

A slightly more complicated strategy which has been suggested is of choosing measurements from as broad a spectrum as possible so that special effects of various Times of Testing or Times of Birth will cancel out each other. This technique might be feasible if a great many Times of Testing or Cohorts could be chosen randomly. Since such special conditions would be difficult to obtain, this technique remains a makeshift used in the absence of a better one. When cultural features which might produce a Time of Testing effect are very broad, i.e., increased use of drugs in the sixties, practicality of the method would be further reduced.

Schaie (1965) suggests a technique for minimizing Time of Testing effects which is in effect the reverse of the one just discussed. If Age and Cohort (Time of Birth) measurements are increased systematically, and Times of Testing held to a minimum, Time of Testing confounding would become trivial by comparison. To determine net Cohort differences he suggests:

The investigator who is interested in the determination of net time or cohort differences . . . can always obtain such information by analyzing the age range or age ranges at which

little or no maturational change occurs. Practically all organisms show such periods for most variables of interest.

Kessen (1965) recommends that measures pertinent to cultural shift be included in the design of the longitudinal project. This strategy may ignore important variables or may make the project cumbersome.

Nesselroade (1970) reports a technique which he considers promising. He says:

Tucker (1963) has generalized the basic factor analytic model to simultaneously handle data involving more than two modes of classification. . . . Although the algebra involved in explicating the structure of the model is demanding, the possibilities for developmental psychology become obvious when one considers that one mode or classification dimension may be time of measurement.

From a different point of view Wohlwill (1970a) does not consider age an independent variable. With its two related attributes, cohort differences and time of testing, he considers it an extraneous dependent variable operating systematically with a great many variables of interest in developmental psychology. As such it affects results in varying degrees depending on many factors.

Cultural Shift in Developmental Psychology Itself.

Radical changes in developmental psychology itself in the past 50 years affected profoundly the early long-range longitudinal studies. Baldwin (1960) describes the quandary in which the far-sighted researchers of the 20's found themselves as their studies progressed. Biological orientation of the twenties shifted to testing of developmental theories of personality and for some child psychologists a preoccupation with mental testing. More recently many child psychologists find

experimental and behavioristic psychology more congenial orientations.

Not only were there changes in theoretical frames of reference in developmental psychology, but experimental techniques known only to a few at the beginning of the century became popular and new applications were developed. No direct remedy is available for the researcher who finds a new and better measuring device than the one he is using after the project is well advanced.

Summary of Cultural Shift.

The problem of cultural shift is a crucial one in longitudinal study. It is essentially the problem of measuring an on-going process. The often-repeated expression of the need for a mathematics capable of dealing with multidimensional problems of change is particularly apt in this connection. As it becomes more apparent that few variables of interest in psychology are static, both the method and the problem will hopefully be given more creative attention.

After considering the problem of cultural change related to measurement of intelligence and suggesting several remedies Kuhlen and Thompson (1970) suggest that a new conception and definition of intelligence may be in order. Not only may conceptualizations of subject matter need to be changed but also the role of the individual study in answering our questions. Perhaps the rapprochement of methods and statistical techniques advocated by Cronbach (1957) is the best answer now available to the problem of cultural shift.

III. SAMPLE SELECTION, RANDOMIZATION, AND GENERALIZABILITY OF THE STUDY

The Basic Dilemma.

Sample selection in longitudinal study presents a basic dilemma even more difficult to resolve than the cultural shift paradox. As mentioned in the previous chapter in the experimental study both experimental and control groups are randomly selected from the same population. The groups are differentiated by manipulation of the independent variable.

On the other hand in order to preserve the "natural" situation in longitudinal study, experimental controls and manipulation of the independent variable are avoided. The experimental group differs from the control group in that the experimental group is selected on the basis of an independent variable (Reinert 1970), i.e., "intelligence" in the Terman study. Some groups are selected in this way in experimental study particularly when analysis of variance designs are used, but in longitudinal study this problem is unavoidable. We can easily see that when the experimental group is selected in this way, randomization is limited.

When randomization is impaired, the statistical validity of results is impaired and generalizability of the study to other groups is thrown in doubt. Thus, to accomplish our purpose, to study a natural situation over time, we seem to have set the stage to defeat the purpose of the study in the end by limiting generalizability of results and calling into question measures of probability.

Kodlin and Thompson (1958) point out the obvious: "It should be apparent that, whatever the method of study, interest in the application of findings will not be limited to the group furnishing the observations." Of course all experiments have limited generalizability. Campbell and Stanley (1969) treat this problem extensively. They maintain that "external validity [generalizability] , like the question of inductive inference is never completely answerable." Both longitudinal and experimental researchers will find much of interest in their designs. They point out generalization from the experimental setting to natural settings is often a problem for the experimental researcher. In longitudinal study which is conducted in a natural setting, this problem is not pertinent of course. Reinert (1970) also discusses importance of sampling procedures in more detail.

Pretesting. Pretesting is often suggested as a means of determining whether experimental and control groups are comparable on variables other than the independent variable. (Campbell and Stanley 1969) Pretesting has two disadvantages. (1) Possible variables that may be operating systematically with the independent variable may not be included in the tests. This objection may be countered with Campbell and Stanley's (1969) often repeated principle of "eliminating plausible alternative hypotheses" wherever possible. Pretesting does eliminate at least some rival hypotheses. (2) Pretesting may sensitize the subject to the variables to be tested. (Lana 1969) Pretest sensitization is a well documented phenomenon and need not be dealt with at length here, but it is a factor to be studied carefully considering that this artifact may plunge results of the study into

even more serious doubt than the "rival hypotheses" the pretests were designed to eliminate. (Campbell and Stanley 1969) Campbell and Stanley suggest two designs which use pretesting rather than randomization.

Matching. Matching is often considered a remedy in a number of experimental circumstances. In this instance, Campbell and Stanley (1969) find it an ineffective cure. (1) The sample may be greatly reduced when matching is used. (2) All dependent variables operating systematically with the independent variable may not be matched so that the groups are still not comparable in that sense. (3) Matching may remove extreme cases of the experimental group thus affecting measurements of the independent variable on the basis of which the group was originally selected. (4) On the other hand, when the experimental group is held intact and the control group matched to it, serious regression effects may be observed. Campbell and Stanley (1969) cite an example in a study of psychotherapy effectiveness. In this study the independent variable was application for therapy and the dependent variable was "dissatisfaction with one's own personality." Campbell and Stanley describe this problem:

Suppose the experimental group consists of therapy applicants and the matched control group of "normal" persons. Then the control group will turn out to represent extreme low scores from the normal group (selected because of their extremity), will regress on the posttest in the direction of the normal group average, and thus will make it less likely that a significant effect of therapy can be shown, rather than produce a spurious impression of efficacy for the therapeutic procedure.

Campbell and Stanley point out covariance analysis and other techniques which are effective in some instances for testing effects

of the experimental variable without the procedure of matching.

The Fortuitous Circumstance. In addition to pretesting the researcher may be alert for other fortuitous measures of the groups which may be available. Campbell and Stanley (1969) point out an ingenious method of avoiding randomization in an experiment.

. . . for a design lacking a pretest . . . to occur naturally requires very special circumstances, which almost never happen. Even so, in keeping with our general emphasis upon the opportunistic exploitation of those settings which happen to provide interpretable data, one should keep his eyes open for them. Such settings will be those in which it seems plausible that exposure to X was lawless, arbitrary, uncorrelated with prior conditions. Ideally these arbitrary exposure decisions will also be numerous and mutually independent.

In the example which they cite, turn-signaling on the part of the car ahead is the independent variable; the presence or absence of turn-signaling by the following car is the dependent variable. This completely fortuitous circumstance was used to demonstrate "a significant imitation, modeling, or conformity effect in agreement with many laboratory studies."

Selection of Control Groups.

As noted earlier in this section in experimental study both experimental and control groups are as nearly alike as possible. A control group is randomly selected from the same population as the experimental group. In longitudinal research the experimental group is selected on the basis of some independent variable. In selection of a control group the problem is to find a group which is comparable to the experimental group except in the independent variable. In a study of delinquents, for instance, this problem may be solved as

easily as pulling folders randomly from records of schools the delinquents attend.

The problem of finding suitable controls is not always an easy one to solve. The circumstance described above may require permission on several levels before records are available. The many studies which have found suitable groups attest that the problem is not insoluble.

Problem of Repeated Measurements.

When the design of a study, either experimental or longitudinal, includes repeated measurements, a problem of lack of independence of measurements is introduced. The problem of repeated measurements is well-documented in the literature and will not be considered here in detail. The chief panacea is in planning sufficient control groups. When the problem of finding suitable controls is solved in longitudinal study, the problem of repeated measurements is routine.

Miscellaneous Problems of Sampling.

Problems of sampling, i.e., the "volunteer subject," and other sources of bias of the original sample, are important in planning the longitudinal study. Like the problem of suitable control groups they are common to all research and are well documented. (Rosenthal and Rosnow 1969)

When possible number of subjects is limited, the researcher may unnecessarily hesitate to continue with the study. When results of a number of studies even using a limited number of subjects point in the same direction, however, hypotheses are strengthened. (Campbell and Stanley 1969)

Kodlin and Thompson (1958) present methods of determining optimal sample size for precision desired.

Summary of Problems Related to Sample Selection.

Although the dilemma presented by selection of groups in longitudinal study is a serious one, a variety of strategies have been discovered for resolving it. Randomization is perhaps the surest but not the only way of enhancing generalizability of the study. The researcher may find it a long wait to discover a natural situation of the type described by Campbell and Stanley (1969) or an independent variable which is known to be randomly distributed in the total population. In the last analysis he can describe the sample he has as carefully as possible and the reader can judge whether the sample is pertinent to his own problems. The researcher may elect to do this in any event.

IV. ATTRITION

Sources of limited generalizability of longitudinal research have been shown relative to cultural shift, sample selection, and randomization. After the sample is selected, generalizability is affected further by attrition in the study over time both longitudinal and experimental.

Attrition in longitudinal research is depletion of the sample from whatever source. According to Kessen (1960) "Whatever technique is used to enlist a sample, there will almost invariably be some subjects who refuse to participate or who drop out of the study later on."

Attrition a Soluble Problem.

Attrition has long been considered an insoluble problem in longitudinal research. (Schaie 1965) Of course for the "life-span" researcher death of subjects as well as the exceptionally long sequence present a problem not pertinent for most developmental research.

Several researchers have shown great perseverance in keeping attrition rates low even over long periods. Their studies demonstrate that attrition problems are soluble. Although Robins (1966) worked from records over 30 years old, she was able to locate 90 per cent of her subjects, 88 per cent experimental and 98 per cent of controls. Bayley (1968) was able to locate all except 2 of her 63 subjects after 33 years.

Robins describes her follow-up methods in some detail. Her first procedure in locating subjects was to submit identifying information to the Social Service Exchange which was still in existence in St. Louis at that time. She also contacted social agencies and former neighbors located by looking up last addresses in the city directory or telephone directory. She usually found either the subject or someone who knew the subject's address. She located 70 per cent of the subjects in this way.

She and her workers then searched police records, Veteran's Administration records, Selective Service records, probation and parole records, former employers, alumni records of colleges, files of commercial credit companies, driver's license registrations and letters forwarded by the Social Security Administration. Long distance calls to anyone bearing the last name or the mother's maiden name were effective in locating relatives in small towns. Driver's license registrations

proved most fruitful of these methods.

Of 548 subjects not known to be dead 76 per cent were interviewed personally. Only 54 refused to be interviewed. Refusals were asked again to participate. Small sums of money were offered; a physician was offered instead of a lay interviewer; a male interviewer was offered if a female had been refused. Twenty refusals were won over in this way. Local subjects refused more often than those at a distance. When the interviewer came from a distance, more importance seemed to be attached to the study.

The high level of compliance may have been due in part to pretesting schedules and interview techniques. Interviewers found best results when they appeared at the door with credentials ready to begin the interview in the initial contact. Letters were found to be one of the poorest contacts. Questions were pretested for comprehension and productiveness of information, and the questionnaire was revised twice.

Droege (1971) found " a variety of techniques is required when detailed information must be obtained . . ." He affirms "nearly 100% follow-up success is possible if detailed data are not required." Droege found successful follow-ups affected by amount of information desired, length of time since contact, and experience of interviewers. Letters were his primary resource and the cheapest technique used. The most effective techniques, however, were personal visits, telephone calls, information from local employment service offices, and contacts with individuals in the community.

Bias of the Sample due to Attrition.

Attrition does not affect generalizability of the study if effects are random, but this is seldom the case. Attrition generally introduces a bias, often of variables of interest or perhaps especially of variables of interest. Kodlin and Thompson (1958) point out: "Such bias will operate if the trait under study is in any way related to factors affecting the loss, a relationship not necessarily intelligible to the investigator even if he knows the overt reason for this 'loss.'" They add: ". . . The longitudinal approach has the advantage of providing at least a partial check on bias since for the lost group some information will be available, gathered before the loss occurred.

Bayley's study (1968) illustrates this advantage of the longitudinal study which the cross-sectional study does not have. By planning in advance some her data indicated relevant characteristics of those who have dropped out or disappeared. In cross-sectional study, as Kuhlen (1951) points out, when measuring differences, i.e., between first and fifth grade subjects, the researcher cannot know in what way the fifth grade sample has been affected by dropouts. He can only know that the older grade will represent a different section of the school population than the younger grade.

Bayley's study showed that the socio-economic distribution of her sample remained essentially unchanged. "This scale is a combination of their parent's education, occupation, income, and a rating of the home and neighborhood made at the start of the study. . ." In longitudinal study if the study becomes biased, the researcher can report how much and in what direction the bias has occurred. According to Kodlin

and Thompson: "Bias can be assumed if the 'losses' differ already from 'remainders' prior to loss." Thus, the reader may know extent to which bias has occurred and in what way it may have affected generalizability of the study.

Bias when Attrition is not Controlled.

If little or no attempt has been made to keep attrition rates low, much research shows that attrition is an important source of bias of the sample. In a study by Baltes, Schaie and Nardi (1971) only 60.4 per cent of the subjects were located after only 7 years, a very high rate of attrition especially inasmuch as many of the subjects were relatively young. This study was originally a study of cognitive behavior.

Analysis of the attrited sample shows:

. . . at the initial occasion of measurement . . . , those subjects who could be retested after 7 years, show on the average higher general intelligence, more cognitive flexibility, and more visuo-motor flexibility when compared with those subjects who could not be retested after 7 years.

Ames and Walker (1965) report that their sample of late adolescents also became progressively biased in a positive direction due to school dropouts. Boys who continued were more productive, more sensitive and more tense. Anderson and Cohen (1939) demonstrate radical effect of attrition on curves. When all data with incomplete dental arch series were included, curves became very irregular and altogether unlike expected relatively smooth curves.

In all three of these cases results are affected in various ways and population to which the study might be generalized is left in doubt.

When Sample Size is Reduced by Experimenter.

In the event that the investigator himself decides to decrease the sample according to some predetermined scheme Kodlin and Thompson (1958) said:

. . . If the reduction in the number of the individuals observed in the second time point is imposed by the investigator, we may assume that he is aware of the mechanism by which he performed this reduction and as a result has some idea of the consequences . . .

They review in mathematical detail these "consequences" in terms of precision of results when the sample size is deliberately manipulated by the experimenter. They distinguish several categories of replacement and discuss effects of matching.

Mobility and Attrition.

Mobility is often cited as a chief source of attrition. Baltes, Schaie, and Nardi (1971) in the study cited previously point out:

. . . Comparison of socioeconomic data for the original and attrited sample suggests that the attrition was not significantly biased with respect to demographic variables, except for indexes of mobility.

Douglas and Blomfield (1956) evaluated effect of mobility on the sample in a study of reliability of longitudinal surveys. They concluded that the sample would have become biased due to mobility of subjects if the survey had not been nationwide (England). Robins (1966) attributes much of the success of her project to having funds for travel and interviewers to trace subjects who have moved.

Better documentation of mobility as a factor in longitudinal study is much needed. Douglas and Blomfield found, for instance, that mobility within the community showed downward mobility in the class structure

whereas a move outside the community showed ascending class mobility. If repeated study showed principles such as this one to be stable and predictable, the researcher may have gained an important tool in evaluating effects of attrition.

In the United States repeated studies have shown that a substantial number of subjects will have moved, and the researcher may have difficulty finding them if the sample is not contacted for some time. Moran and Swartz (1970) in a longitudinal study of cognitive dictionaries excluded children of students at the University of Texas, the military, and of legislators in Austin, Texas. They felt these categories were likely to move away before the study was completed.

Summary of Attrition. When the sample is altered, unless intentionally by the investigator, change estimates are affected, and generalizability of the study left in doubt. Present methods of reducing attrition are varied and effective. Attrition need not invalidate longitudinal study if it is anticipated and suitable plans are made in advance. Accurate reporting of attrition as with cultural shift and sample selection is the minimum precaution the researcher may take in validating his study.

In this chapter a number of problems of planning the study, cultural shift, and sampling, as well as attrition have been discussed. Chapter IV will summarize trends in modern longitudinal study in terms of some of these problems.

CHAPTER IV

REVIEW OF RECENT STUDIES AND SUMMARY

Review of Modern Longitudinal Studies.

Although a number of modern studies have been used to illustrate problems in Chapter III, this chapter will survey more generally some recent longitudinal studies. Review of recent developmental literature in writing this thesis reveals Jones and Bayley's (1950) statement over 20 years ago quoted previously to be still pertinent today:

". . . psychologists are as yet making little systematic use of the longitudinal approach." Hartup (1968) and White (1969) note the lack of much needed longitudinal data in their fields. The number of studies reported in any particular year does not necessarily reflect the amount of on-going longitudinal activity of course.

A search of The Psychological Abstracts under Developmental Psychology shows cross-sectional studies to be in great preponderance. A number of experimental studies were noted which mark a trend in that direction. Several studies designated as "longitudinal studies" are experimental studies over time.

Some recent studies have been noted which appear to be longitudinal studies but are apparently studies from records. Often these are records from "old" longitudinal studies which are reanalyzed or to use Sontag's (1971) expression, mined. These studies together with some of the life-span developmental studies were very long in comparison with other

modern studies of more limited lineage and objectives. This variability together with the inherent individuality of each longitudinal study makes generalizations about "modern" studies difficult.

Almost all of the modern studies reviewed for this thesis had control or comparison groups. Numbers of subjects were usually sufficient for purposes of computing probabilities. Attrition rate was commonly reported although not well controlled in all cases. Of the problem areas considered in Chapter III cultural shift was less often specifically considered. In some instances of course it was not relevant. By and large these studies showed a vast methodological improvement over the studies reported by Stone and Onqué in 1959.

A Modern Study Analyzed by Problem Areas.

A follow-up study by Thomas Aschenbach (1971) may be considered an example of "the new breed" of longitudinal study, short (2 years) and with limited objectives.

Objectives of the Study. The objectives of the study are clearly defined:

The purposes of the present study were to examine over a 2-year period achievement and ability measures for the children classified as high or low D scorers when they were in fifth or sixth grade, as well as the stability of the D score classification and the correlations of error scores over two testings 18 months apart.

Specific null hypotheses are not stated. This study is well-oriented to other studies and literature in the field. The measure used had been cross-validated and revised using three previous groups of fifth and sixth graders.

Selection of Variables and Groups. In this study 262 fifth graders and 251 sixth graders were administered the Children's Associative Responding Test (CART). In addition to grade the groups were further divided on the basis of whether the child responded to analogies by association (referred to in the study as a high D responder) or reason (low D responder).

In the follow-up these children were administered the CART again in the seventh and eighth grades respectively. The seventh and eighth graders were described according to ethnic background, socioeconomic status and type of housing and rated on Hollingshead's 7-point scale of occupational status to facilitate generalization of results to other groups. These figures were apparently for the grades as a whole. The Hollingshead SES scores were significantly different for low and high D groups in the fifth grade but not in the sixth. These were the only figures of this type in the study to show cultural differences between high and low D responders.

In the seventh and eighth grades in addition to scores from the readministration of the CART several other evaluations served as dependent variables: 7th grade Otis IQ, current MA in months, SRA achievement grade level, grade point averages for the previous year, grade point averages for the first three months of the current year, and Hollingshead SES scores for father's occupation mentioned above.

Statistics Used. Mean scores and standard deviations were listed in a table according to grade and high or low D score. Sex was added as an independent variable and sex x D score ANOVAS were computed in addition to other analysis of variance.

Table 2 presented a different kind of information in the form of correlations. Initial IQ, MA, and CART total errors of fifth and sixth graders were correlated with current achievement, grade averages, and previous year's grade averages for low and high D scorers.

To test for a possible confounding of D score with IQ, matching was attempted between a high D scorer and a low D scorer of the same sex, similar fifth-grade IQs, and similar MAs. The pattern of correlations was similar to correlations in Table 2.

Repeated measurement ANOVAS for subjects having both IQ scores showed a significant divergence of IQ favoring the low D scorers in each grade. For the seventh graders this difference was accounted for by a rise in the IQs of the low D scorers. The eighth graders showed a similar but not significant trend. This finding seems to represent the dropping out of less able students often noted in both cross-sectional and longitudinal studies.

An "alternative plausible hypothesis" was suggested to account for some other small divergences. Since matching had required selecting low D scorers with unrepresentatively low IQ and high D scorers with unrepresentatively high IQs, Aschenbach considered the possibility that regression effects (discussed in Chapter III) may have contributed to the divergences. When such an effect created by matching is operating, it can have serious consequences in results from data for matched pairs. Aschenbach felt, however, that since the data did not consistently show this effect, results had probably been only negligibly affected by it.

Attrition. Attrition is summarized here: Low D, fifth grade, 6 per cent; high D, fifth grade, 16 per cent; Low D, sixth grade, 14 per cent; high D, sixth grade, 14 per cent. The significance of this attrition was in part assessed by comparing mean IQ scores assessed in fifth grade of the attrited group of low D scorers with those of the original group of low D scorers. The difference in IQ scores was only about 2 points. The other three classifications were similarly assessed with a difference in mean IQ scores of only about .3 points.

Cultural Shift. Although Aschenbach does not specifically mention cultural shift, he is aware of this problem. He has noted the exact date of the testing of 7th and 8th grades. He notes, further, in one instance that a discrepancy of scores may have been due to a time of testing difference. A previous test had been used earlier in the school year than a later test.

He does not point out that differences between grades 5 and 6 and grades 7 and 8 may have been in part due to cohort differences, often a possibility whenever more than one age group is used. In view of the small differences between the groups, however, this possibility is probably negligible. Use of two grades in this way, it is interesting to note, has actually introduced a cross-sectional study concurrently with the longitudinal study.

Aschenbach describes two unfortunate features of the study which are attributable to cultural shift. In the first place, the second administration of the CART contained nine revised items which had not been included in the first administration. Although correlations

between error scores on the initial and current administrations of the test indicated high test-retest reliability, these changes introduced some doubt in interpreting specific error scores for seventh and eighth graders.

Changes in the measuring instrument of this type are a common problem of the longitudinal researcher. Aschenbach could have avoided this difficulty easily by administering the original version of the instrument for the second testing even though he believed the new version to be better. If feasible, he might in addition have administered the nine revised items to these children. If this were not feasible, he would better have tested the new instrument on an entirely new group of children than risk untoward effect in the study as a whole.

The second instance of cultural shift which affected this study was that in the current school year the grading system had been changed from a 5-point system to a 3-point system. Aschenbach made amends for this unavoidable difficulty by including grades from the previous year as well as for the current year.

Results and Conclusions. Aschenbach's discussion of results were far-ranging and a great many tests of significance were made. Since specific null hypotheses had not been made in advance, these tests of significance are subject to some doubt. This criticism might have been avoided in the planning stage of the project by making null hypotheses beforehand.

Objectives of the study were carried out. Achievement and ability measures for the children classified as high or low D scorers when they were in fifth or sixth grade were examined and compared with scores in

seventh and eighth grades. The D score classification was found to be stable over the 18 month period from first to second measurement. The correlations of error scores were also found to be stable. According to Aschenbach (1971):

A primary conclusion that can be drawn from the present study is that the tendency to rely excessively on associative responding is a stable one which, at any given level of ability, adversely affects scholastic performance over at least two school grades. An equally important conclusion is that the CART's psychometric properties make it not only a useful instrument for identifying associative responding, but also an alternative to the group IQ test as a predictor of performance during the early formal operational period of development.

This study is not introduced here as "typical" of modern studies. It does seem to represent the direction of modern longitudinal studies and the increasing problem orientation of modern researchers. The problems demonstrated in this study may appear exaggerated in this analysis. For the most part the problems were minor, the researcher was aware of them, and produced considerable evidence that they had not affected results a great deal.

Another briefer analysis of a recent study by Faterson and Witkin (1970) is shown in the appendix of this thesis. In this schedule problem areas are pin-pointed and assessed with plus and minus values. With regard to the main problem areas the two chief questions implicit in the evaluation are: (1) Have results of the study been affected by the problem? and (2) Is generalizability of the study affected by the problem? This schedule is designed to serve as a checklist for the researcher or reader in evaluating a longitudinal study.

Perils of the Problem-oriented Paper.

Kodlin and Thompson (1958) and Campbell and Stanley (1969) both note that when problems of a method of study are considered in a forthright way, potential researchers may become discouraged and cease to pursue the study further. Although this thesis is problem-oriented, hopefully the effect will be to raise the level of sophistication in longitudinal study rather than to discourage those who might engage in a study over time. The point of view of this paper is well expressed by Roy Mefford (1966) in the following quotation:

It would be foolhardy to minimize the statistical problems as well as the great practical difficulty in the conduct of longitudinal studies. We believe, however, that these can be solved, and that it is now time to accumulate data that will permit their solution.

The researcher may take some comfort in the thought that only a rare study would encompass all of the problems described in this thesis. As Campbell and Stanley demonstrated in the study of turn-signaling cited previously, under especially fortuitous circumstances none of these problems may be involved. A review of modern studies has shown repeatedly that a particular problem may not be relevant to a particular natural research setting or research objective. In a study of intellectual decline as an indicator of pathology in old age, by Blum and Jarvik (1969), death of subjects seemed to create a high attrition rate. These researchers made time of death itself a dependent variable, however, and attrition became a part of the study rather than a problematic side-effect.

Unlike the experimental study where a situation is created according to a prescribed pattern, each natural situation is unique. Whether

a problem is pertinent to that particular situation remains a matter of individual judgement. Although Chapter III describes some techniques of solving frequent longitudinal problems, the researcher may find special circumstances of the study itself dictate problem solutions. Mefford (1966) further optimistically points out, as understanding of the method increases, what appear to be liabilities at this time, may turn out to be assets.

Summary.

This thesis has differentiated traditional longitudinal study from other methods of study: the cross-sectional study, the time-lag study, the experimental study, the retrospective study, and the study from records. Each method of study has unique procedures and problems.

The traditional longitudinal study has been defined as the study: (1) of individual change under natural conditions not controlled by the experimenter, (2) which proceeds over time from the present to the future by measuring the same individuals repeatedly, and (3) which retains individuality of data in analyses.

Some of the main ideas about longitudinal study presented in this thesis are : (1) The longitudinal method is useful. (2) In order to answer some questions which are important in developmental psychology it is necessary. (3) Longitudinal study continues to be used. It cannot be ignored. We have no alternative but to make every effort to understand the nature of the problems of longitudinal study and to try to cope with them.

Longitudinal method has some advantages. It preserves individuality

of data over time when used in conjunction with correlational analysis. Results show covariation of variables and are predictive.

Longitudinal method has some disadvantages. It does not always make a strong "causal" statement. Causality can only be shown when "plausible alternative hypotheses" are eliminated. (Campbell and Stanley 1969) Although showing causality is not the only worthwhile objective of a scientific study, an effort to eliminate other hypotheses strengthens the study.

In this thesis some problem areas of longitudinal study which are either unique to this method or especially difficult have been delineated. Problems related to planning the study were discussed. Cultural shift and attrition were especially emphasized. The dilemma was examined which is posed by sample selection with its related problems of randomization of generalizability of the study, together with the problems of repeated measurements and selection of control groups.

Not only are these problems delineated but considerable evidence is shown that we have already started "to accumulate data that will permit their solution." (Mefford 1966). This paper has presented a number of studies which have considered these problems separately or as a side issue of a study on some other topic. Studies of attrition, i.e., Robins (1966) and Droege (1971), contain much valuable information which may be gathered and made the subject of further research. The excellent work begun by Kuhlen (1940 1951) on cultural shift has been continued by Schaie (1965) and others. More careful analyses of the nature of process are needed in this area. Exploration of

mathematical bases of the method so ably begun in the classic work by Kodlin and Thompson (1958) need to be extended, and statistical methods improved. Almost every longitudinal study has made some contribution to knowledge of the method, if only by pointing out its perplexities.

In this thesis perhaps the most general recommendation for coping with problems of the longitudinal study has been to limit the time and objectives of the study. Limitation of the study together with greater familiarity with problems and procedures of longitudinal method leads to more meaningful studies. When objectives of the study are well defined and time is limited to only enough for specified change to take place, many of the problems of longitudinal study are reduced to manageable proportions. As researchers gain confidence in this tool and knowledge of what it can and cannot do, more longitudinal studies may be undertaken and much needed information may be obtained from natural settings over time.

At the same time that this thesis has stressed differentiation of the longitudinal study from other studies, it has stressed integration of results of longitudinal studies with results of other studies. Without a true federation of the disciplines all our methods can give only wrong answers. (Cronbach 1957) The belief adopted here was expressed by Campbell and Stanley (1969): scientific knowledge is cumulative and not dependent on one crucial experiment.

Baltes (1968) believes "there is a general trend in developmental psychology towards formulating a methodology specifically suited to developmental problems." While this seems to be true, a counter trend seems to be expanding methodological horizons of

developmental psychology to include a wide variety of research methods and statistical techniques. At this time method in developmental psychology is far from crystallized. With further study of the method and research on problems longitudinal study may take a more prominent place in the battery of developmental methods.

A SELECTED BIBLIOGRAPHY

- Ames, L. B. and Walker, R. M. A note on school dropouts in longitudinal research with late adolescents. J. of Genetic Psychology, 1965, 107, 277-279
- Anderson, J. E. Methods of child psychology. In Carmichael, L. (ed.) Manual of Child Psychology (2nd. ed.). N. Y.: Wiley, 1954
- Anderson, J. E., and Cohen, J. T. The effect of including incomplete series in the statistical analysis of longitudinal measurements of children's dental arches. Child Development, 10, 1939
- Andrews, T. G. Methods of Psychology. N. Y.: Wiley and Sons Inc., 1948
- Aschenbach, T. M. The childrens associative responding test: a two-year follow-up. Developmental Psychology, 1971
- Baker, C. T., Sontag, L. W., and Nelson, V. L. Individual and group differences in the longitudinal measurement of change in mental ability. Mental Growth and Personality Development: A Longitudinal Study, Monographs of the Society for Research in Child Development, 1958, 23, No. 2, Serial No. 68
- Baldwin, A. L. The study of child behavior and development. In Mussen, P. H. (ed.) Handbook of Research Methods in Child Development. N. Y.: Wiley, 1960, 3-35
- Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. Human Development, 11, 1968 145-171
- Baltes, P. B., Nesselroade, J. R., Schaie, K. W., and Labouvie, E. W. On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. Developmental Psychology, 1972, 6, No. 1, 1971
- Baltes, P. B., Schaie, K. W., and Nardi, A. Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. Developmental Psychology, 5, No. 1, 1971
- Bayley, N. Behavioral correlates of mental growth: birth to 36 years. American Psychologist, 1968, 23, No. 1, 1-17
- Bell, R. Q. Convergence: an accelerated longitudinal approach. Child Development, 1953, 24, 145-152

- Bereiter, C. Some persisting dilemmas in the measurement of change. In Harris, C. W. (ed.) Problems in Measuring Change. Madison, Wis.: Univ. of Wis. Press, 1963, 3-20
- Birren, J. E. Principles of research on aging. In J. E. Birren (ed.) Handbook of Aging and the Individual. Chicago: Univ. of Chicago Press, 1959
- Bloom, B. S. Stability and Change in Human Characteristics. N. Y.: Wiley, 1964
- Blum, J. E., and Jarvik, L. F. Variations in intellectual decline as indicators of pathology: a longitudinal twin study. Proceedings, 77th Annual Convention, A.P.A., 1969
- Burton, R. V. Validity of retrospective reports assessed by the multitrait-multimethod analysis, Developmental Psychology Monographs, Part 2, Nov. 1970, 1-15
- Butler, J. M., Rice, L. N., and Wagstaff, A. K. Quantitative Naturalistic Research, Englewood Cliffs, N. J.: Prentice-Hall Inc., 1963
- Campbell, D. T. From description to experimentation: interpreting trends as quasi-experiments. In C. W. Harris (ed.) Problems in Measuring Change. Madison, Wis.: Univ. of Wis. Press 1963
- Campbell, D. T. Prospective: artifact and control. In Rosenthal, R., and Rosnow, R. L. (eds.). N. Y.: Academic Press, 1969
- Campbell, D. T. and Fiske, D. W., Convergent and discriminant validation by the multi-trait multi-method matrix. Psychological Bulletin, 1959, 56, 81-105
- Campbell, D. T. and Stanley, J. C., Experimental and Quasi-Experimental Designs for Research. Chicago: R. McNally, 1969
- Cattell, R. B. The Description and Measurement of Personality. Yonkers, N. Y.: World Book, 1946
- Cattell, R. B. The structuring of change by P-technique and incremental R-technique. In Harris, C. W. (ed.), Problems in Measuring Change. Madison, Wis.: Univ. of Wis. Press, 1963
- Cattell, R. B. Patterns of change: measurement in relation to state dimension, trait change, lability, and process concepts. In R. B. Cattell (ed.) Handbook of Multivariate Experimental Psychology. Chicago: R. McNally, 1966, 355-402
- Cattell, R. B. Handbook of Modern Personality Study. Chicago: Aldine Press, 1970

- Churchill, E. In Falkner, F. (ed.) Human Development. Philadelphia: W. B. Saunders Co., 1966
- Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684
- Damon, A. Discrepancies between findings of longitudinal and cross-sectional studies in adult life: physique and physiology. Human Development, 1965, 8, 16-22
- Davids, A., and Holden, R. H. Consistency of maternal attitudes and personality from pregnancy to eight months following childbirth. Developmental Psychology, 1970, 2, No. 3, 364-366
- Davies, D. F. Mortality and morbidity statistics. I. Limitations of approaches to rates of aging. J. of Gerontology, 1954, 186-195
- Dearborn, W. F., and Rothney, J. W. M. Predicting the Child's Development. Cambridge, Mass.: Sci-Art. Publishers, 1941
- Droege, R. C. Effectiveness of follow-up techniques in large-scale longitudinal research. Developmental Psychology, 5, No. 1, 1971
- Droege, R. C. and Crambert, A. C. Follow-up techniques in a large scale test validation study. J. of Applied Psychology, 1965, 49, 253-256
- Douglas, J. W. B. and Blomfield, J. M. The reliability of longitudinal surveys. Milbank Memorial Fund Quarterly, 1956, 34, 227-252
- Escalona, S. and Heider, G. M. Prediction and Outcome: A Study in Child Development. N. Y.: Basic Books, 1959
- Faterson, H. F., and Witkin, H. A. Longitudinal study of development of the body concept, Development Psychology, 1970, 2, No. 3, 429-438
- Fisher, R. A. Statistical Methods for Research Workers. N. Y.: Hafner Pub. Co., 1958
- Fisher, R. A. The Design of Experiments. N. Y.: Hafner Pub. Co., Reprinted 1966
- Glueck, S. and Glueck, E. Delinquents and Nondelinquents in Perspective. Cambridge: Harvard Univ. Press, 1968
- Goldfarb, N. An Introduction to Longitudinal Statistical Analysis; the Method of Repeated Observations from a Fixed Sample. Glencoe, Ill. Free Press, 1960

- Goulet, L. R. and Baltes, P. B. (eds.) Life-Span Developmental Psychology: Research and Theory. N. Y.: Academic Press, 1970
- Guilford, J. P. Psychometric Methods. (2nd ed.) N. Y.: McGraw-Hill 1954
- Guilford, J. P. Fundamental Statistics in Psychology and Education (4th ed.) N.Y.: McGraw-Hill Book Co. 1965
- Harmon, H. H. Modern Factor Analysis. Chicago: Univ. of Chicago Press, 1967
- Harris, C. W. (ed.) Problems in Measuring Change. Madison, Wis.: Univ. of Wis. Press, 1963
- Hartup, W. W. J. of Research and Development in Education, 1, No. 3, 1968, 16-29
- Hays, W. L. Statistics for Psychologists. N. Y.: Holt, Rinehart and Winston, 1953
- Hoffman, M. L. and Hoffman, L. N. (eds.) Review of Child Development Research. 1, N. Y.: Russell Sage, 1964
- Horn, J. L. Organization of data on life-span development of human abilities. In Goulet, L. R. and Baltes, P. B. (eds.) Life-Span Developmental Psychology: Theory and Research. N. Y.: Academic Press, 1970
- Johnson, P. O. Statistical Methods in Research. N. Y.: Prentice-Hall Inc., 1949
- Jones, H. E. Problems of method in longitudinal research. Vita Humana, 1, 1958, 93-99
- Jones, H. E. and N. Bayley, Growth, development and decline. Annual Review of Psychology, 1, No. 1, 1950
- Kagan, J. American longitudinal research. Child Development, 35, 1964, 1-32
- Kagan, J. and Moss, H. Birth to Maturity, a Study in Psychological Development. N. Y.: Wiley, 1962
- Kessen, W. Research design in the study of developmental problems. In Muss, P. H. (ed.) Handbook of Research Methods in Child Development. N. Y.: Wiley, 1960, 40 ff
- Kerlinger, F. N. Foundations of Behavioral Research. N. Y.: Holt Rinehart, and Winston, Inc., 1964

- Kodlin, D. and Thompson, D. I. An appraisal of the longitudinal approach in the study of growth and development. Monographs of the Society for Research in Child Development, 1958, 23, No. 1, Serial No. 67
- Kuhlen, R. G. Social change: A neglected factor in psychological studies of the life span. School and Society, 1940, 52, 14-16
- Kuhlen, R. G. Psychology of Adolescent Development. N. Y.: Harper and Brothers, 1951
- Kuhlen, R. G. and Thompson, G. G. Psychological Studies of Human Development. (3rd ed.) N. Y.: Appleton-Century-Crofts, 1970
- Lana, R. E. Pretest sensitization. In Rosenthal, R. and Rosnow, R. L. (eds.) Artifact in Behavior Research. N. Y.: Academic Press 1969
- Lord, F. M. Elementary models for measuring change. In Harris, C. W. (ed.) Madison, Wis.: Univ. of Wis. Press, 1963
- Macfarlane, J. W. Studies in child guidance I. Methodology of data collection and organization. Monographs of the Society for Research in Child Development, 1938, 3, No. 6, Serial No. 19
- Mainland, D. Elementary Medical Statistics. Philadelphia: W. B. Saunders, 1952
- McCormick, C. C. and Schnobrich, J. N. IES Arrow-Dot Performance in Two Montessori Schools. Perceptual and Motor Skills, 1969, 29, 235-239
- Mefford, R. B. Structuring physical correlates of mental processes and states: the study of biological correlates of mental processes. In Cattell, R. B. (ed.) Handbook of Multivariate Experimental Psychology. Chicago: R. McNally, 1966
- Merrill, M. The relationship of individual growth to average growth. Human Biology, 1931, 3, 37-70
- Moran, L. J., and Swartz, J. D. Longitudinal study of cognitive dictionaries from ages nine to seventeen. Developmental Psychology, 1970, 3, No. 1, 21-28
- Mussen, P. H., Handbook of Research Methods in Child Development. N.Y.: Wiley, 1960
- Nesselroade, J. R. Application of multivariate strategies to problems of measuring and structuring long-term change. In Goulet, L. R. and Baltes, P. B. (eds.) Life-Span Developmental Psychology: Research and Theory. N. Y.: Academic Press, 1970

- Newsletter, American Psychological Assn., Division on Developmental Psychology, 1968, 1-3
- Page, E. B. and Pinango, E. Information gained through combining psychometric studies. Proceedings American Psychological Assn. 1970, 5, Pt. 1, 127-128
- Platt, J. R., Strong inference. Science, 1964, 146, 347-353
- Prospectus on Research, White House Conference on Children and Youth. Washington, D. C.: U. S. Government Printing Office, Supt. of Documents, 1960
- Rees, A. H. and Palmer, F. H. Factors related to change in mental test performance, Developmental Psychology Monographs, Sep. 1970, 3, No. 2, Part 2,
- Reese, H. W. and Lipsitt, L. P. Experimental Child Psychology. N.Y.: Academic Press, 1970
- Research Relating to Children, bulletin, Washington, D. C.: Supt. of Documents, U. S. Government Printing Office
- Riegel, K. F., Riegel, R. M., and Meyer, G. A. A study of the drop-out rates in longitudinal research on aging and the prediction of death. J. of Personality and Social Psychology, 1967, 4, 342-348
- Riegel, K. F., Riegel, R. M., and Meyer, G. A. The prediction of retest resisters in research on aging. J. of Gerontology, 1968, 23, 370-374
- Robins, L. N. Deviant Children Grown Up. Baltimore: Williams and Wilkens, 1966
- Rose, C. L. Representativeness of volunteer subjects in a longitudinal aging study. Human Development, 1965, 8, 152-156
- Rosenthal, R. and Rosnow, R. L. (eds.) The volunteer subject. In Artifact in Behavior Research. N.Y.: Academic Press, 1969
- Schaie, K. W. Cross-sectional methods in the study of psychological aspects of aging. J. of Gerontology, 1959, 14, 208-215
- Schaie, K. W. A general model for the study of developmental problems, Psychological Bulletin, 1965, 64, 92-107
- Tanner, J. M. A guide to American growth studies. Yearbook Physical Anthropology, 1947, 3, 230-238
- Terman, L. M., Baldwin, B. T., and others, Genetic Studies of Genius Vol. 1, Mental and physical traits of a thousand gifted children Stanford Univ., Calif.: Stanford Univ. Press, 1926

- Terman, L. M., and Oden, M. H. The Gifted Group at Mid-Life. Stanford: Stanford Univ. Press, 1959
- Thomas, A., et al. Behavioral Individuality in Early Childhood. N.Y.: N.Y. Univ. Press, 1963
- Thompson, G. G. Children's groups. In Mussen P. H. (ed.) Handbook of Research Methods in Child Development. N.Y.: Wiley, 1960
821-853
- Thorndike, E. L. Prediction of Vocational Success. N.Y.: The Commonwealth Fund, 1934
- Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In Harris, C. W. (ed.) Madison, Wis.: Univ. of Wis. Press, 1963
- Turner, M. B. Philosophy and Science of Behavior. N.Y.: Appleton-Century-Crofts, 1965
- Walkey, F. H. and Boshier, R. Changes in semantic differential responses over two years. Psychological Reports, 1969, 24, No. 3, 1008-1010
- Welford, A. T. Ageing and Human Skill. London: Oxford Univ. Press, 1958
- White, B. L. Child development research: an edifice without a foundation. Merrill-Palmer Quarterly, 1969, 15, No. 1, 49-79
- Willems, E. P. and Raush, H. L. (eds.) Naturalistic Viewpoints in Psychological Research. N.Y.: Holt, Rinehart, and Winston, 1969
- Wohlwill, J. F. Methodology and research strategy in the study of developmental change. In Goulet, L. R. and Baltes, P. B. (eds.) Life-Span Developmental Psychology: Research and Theory. N.Y.: Academic Press, 1970
- Yarrow, M. R., Campbell, J. D., and Burton, R. V. Recollections of childhood: a study of the retrospective method, Monographs of the Society for Research in Child Development, 1970, 35, No. 5, Serial No. 138
- Yule, G. U., Why we sometimes get nonsense correlations between time series. J. Roy. Statist. Ass., 1926, 89, 1-64

APPENDIX

SCHEDULE FOR EVALUATION OF A LONGITUDINAL STUDY*

1. How long was the study in months? 168 months
2. How many subjects were there? 30 boys, 30 girls, experimental
26 boys, 27 girls, control
N = 113
3. Are objectives of the study clearly defined? -
4. Are results of the study clearly related to stated objectives
of the study? -
5. Have "plausible alternative hypotheses" been eliminated? -
6. Could objectives of the study have been satisfied only by this
method of study? (Is only an average wanted? Are successive
observations needed? Is an early observation related to a
later observation?) +
7. Are variables under study clearly defined? +
8. Do variables operate freely without undue interference of the
study itself on the subject? ? (Not enough information)
9. Were control or comparison groups used? +

A plus value in the following problem areas means that results of the study have not been affected by that problem; a negative value means that results have been affected by that problem:

10. Selection of groups? (Pretest sensitization, lack of randomization, matching, regression effects due to selection of sample from extreme scores, etc.) +

11. Attrition? -
12. Cultural shift? +
13. Repeated measurements? +

A plus value in the following problem areas means that generalizability of results to other groups has not been affected by that problem; a negative value means generalizability has been affected by that problem:

14. Selection of groups? +
15. Attrition? -
16. Cultural shift? +

*Data and plus and minus values in the schedule apply to a study by Faterson and Witkin, (1970) "Longitudinal Study of Development of Body Concept."